U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-283 448

# Use of Meteorological Data in Air Quality Trend Analysis

Technology Service Corp, Santa Monica, Calif

Prepared for

Environmental Protection Agency, Research Triangle Park, N C Office of Air Quality Planning and Standards

May 78

United States
Environmental Protection
Agency

Office of Air Quality
Planning and Standards
Research Triangle Park NC 27711

EPA-450/3-78-024
May 1978

Air

EPA

PB 283 448

# Use of Meteorological Data in Air Quality Trend Analysis

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

| TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing) | | |
|---|---|---|
| 1. REPORT NO. EPA-450/3-78-024 | 2. | 3. RECIPIENT'S ACCESSION NO. PB283448 |
| 4. TITLE AND SUBTITLE Use of Meteorological Data in Air Quality Trend Analysis | | 5. REPORT DATE May 1978 |
| | | 6. PERFORMING ORGANIZATION CODE |
| 7. AUTHOR(S) Melvin D. Zeldin and William S. Meisel | | 8. PERFORMING ORGANIZATION REPORT NO. |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Technology Service Corporation 2811 Wilshire Blvd. Santa Monica, California 90403 | | 10. PROGRAM ELEMENT NO. |
| | | 11. CONTRACT/GRANT NO. 68-02-2318 |
| 12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air and Waste Management Office of Air Quality Planning and Standards Research Triangle Park, North Carolina 90403 | | 13. TYPE OF REPORT AND PERIOD COVERED |
| | | 14. SPONSORING AGENCY CODE 200/04 |
| 15. SUPPLEMENTARY NOTES | | |

16. ABSTRACT

Meteorological influences upon air quality trend analyses can complicate the evaluation of air pollution control efforts. It is important to isolate the meteorological effects in order to determine air quality trends as a function of emissions. This report surveys existing methods for meteorologically adjusting air quality trends, including a review of known relationships between specific air pollutants and various meteorological parameters, and presents a recommended methodology to normalize air quality trends with respect to "typical" year meteorology. Procedures involve the determination of mutually exclusive meteorological classes and the treatment of air quality variables stratified according to the established meteorological classes. A hypothetical example is included to illustrate both the mathematical processes and the interpretation of the methodological results.

| 17. KEY WORDS AND DOCUMENT ANALYSIS | | |
|---|---|---|
| a. DESCRIPTORS | b. IDENTIFIERS/OPEN ENDED TERMS | c. COSATI Field/Group |
| Trend Analysis<br>Air Pollution<br>Meteorology | | |
| 18. DISTRIBUTION STATEMENT Unlimited | 19. SECURITY CLASS (This Report) Unclassified | 21. NO. OF PAGES 101 |
| | 20. SECURITY CLASS (This page) Unclassified | 22. PRICE A06-A01 |

EPA Form 2220-1 (9-73)

EPA-450/3-78-024

# Use of Meteorological Data in Air Quality Trend Analysis

by

Melvin D. Zeldin and William S. Meisel

Technology Service Corporation
2811 Wilshire Boulevard
Santa Monica, California 90403

Contract No. 68-02-2318

EPA Project Officer: Neil H. Frank
Monitoring and Data Analysis Division

Prepared for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

May 1978

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - in limited quantities - from the Library Services Office (MD-35), U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

This report was furnished to the Environmental Protection Agency by Technology Service Corporation, Santa Monica California 90403 in fulfillment of Contract No. 68-02-2318. The contents of this report are reproduced herein as received from Technology Service Corporation. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention of company or product names is not to be considered as an endorsement by the Environmental Protection Agency.

Publication No. EPA-450/3-78-024

## ABSTRACT

Meteorological influences upon air quality trend analyses can complicate the evaluation of air pollution control efforts. It is important to isolate the meteorological effects in order to determine air quality trends as a function of emissions. This report surveys existing methods for meteorologically adjusting air quality trends, including a review of known relationships between specific air pollutants and various meteorological parameters, and presents a recommended methodology to normalize air quality trends with respect to "typical" year meteorology. Procedures involve the determination of mutually exclusive meteorological classes and the treatment of air quality variables stratified according to the established meteorological classes. A hypothetical example is included to illustrate both the mathematical processes and the interpretation of the methodological results.

## PREFACE

This report contains an overview and synthesis of methods which have been proposed for adjusting observed measurements of air quality to consistent meteorology. The recommended approach has not been tested empirically. Our recommendations are thus qualified by the recognition that application of the recommended procedure to measurement data is likely to reveal opportunities for improvement.

TABLE OF CONTENTS


| Section | | | Page |
|---|---|---|---|
| ABSTRACT | | | ii |
| PREFACE | | | iii |
| FIGURES | | | vi |
| TABLES | | | vii |
| ACKNOWLEDGMENTS | | | viii |
| 1.0 | INTRODUCTION | | 1 |
| | 1.1 | SCOPE OF STUDY | 3 |
| | 1.2 | OUTLINE OF RECOMMENDED PROCEDURE | 5 |
| 2.0 | METEOROLOGICAL EFFECTS UPON AIR POLLUTION | | 7 |
| | 2.1 | METEOROLOGICAL VARIABLES AFFECTING AIR POLLUTION | 7 |
| | | 2.1.1 Key Meteorological Variables by Pollutant | 7 |
| | | 2.1.2 Availability of Meteorological Data | 17 |
| | | 2.1.3 Representativeness of Meteorological Data | 18 |
| | 2.2 | VARIATIONS IN AIR QUALITY DATA DUE TO METEOROLOGY | 22 |
| | | 2.2.1 Variation in Meteorological Effect Among Pollutants | 23 |
| | | 2.2.2 Variation in Meteorological Effect Among Air Quality Parameters | 26 |
| 3.0 | APPROACHES TO METEOROLOGICAL ADJUSTMENT OF AIR QUALITY TRENDS | | 29 |
| | 3.1 | POSTULATION FROM BASIC PRINCIPLES | 30 |
| | 3.2 | THE REGRESSION APPROACH | 37 |
| | 3.3 | CLASSIFICATION APPROACHES TO METEOROLOGICAL ADJUSTMENT | 45 |
| | 3.4 | METEOROLOGICALLY INVARIANT STATISTICS | 47 |
| 4.0 | RECOMMENDED PROCEDURES FOR METEOROLOGICAL ADJUSTMENT OF AIR QUALITY | | 49 |
| | 4.1 | SELECTION OF METEOROLOGICAL CLASSES | 49 |
| | | 4.1.1 Definition Through Experience and Expertise. | 52 |
| | | 4.1.2 Empirical Derivation of Meteorological Classes | 53 |
| | 4.2 | A METEOROLOGICAL ADJUSTMENT PROCEDURE | 58 |

TABLE OF CONTENTS (CONT.)


| Section | | Page |
|---|---|---|
| 4.3 | CHOICE OF THE AIR QUALITY MEASURE TO BE ADJUSTED . . . . | 65 |
| 4.4 | AN EXAMPLE OF METEOROLOGICAL ADJUSTMENT OF AIR QUALITY DATA . . . . . | 68 |
| REFERENCES | . . . . . | 77 |
| APPENDIX A | . . . . . | 81 |
| APPENDIX B | . . . . . | 84 |

## FIGURES


| Number | | Page |
|---|---|---|
| 2.1 | San Bernardino Smog Season: Maximum Oxidant vs. Wind Velocity at Time of Occurrence . . . . . | 15 |
| 3.1 | Annual Value of Pollution Measure Versus Annual Values of Meteorological Index . . . . . | 35 |
| 3.2 | Example of Linear and Non-linear Relationships . . . . . | 43 |
| 4.1 | First Stage of Discrimination Employed by Davidson . . . . . | 55 |
| 4.2 | Full Prediction Scheme by Davidson . . . . . | 56 |
| 4.3 | Decision Scheme of Figure 4.2 Drawn as a Decision Tree . . . | 57 |
| 4.4 | Results of Hypothetical Example Showing Actual $SO_2$ Trends With Associated Meteorologically Adjusted Trends . . . . . | 73 |
| A-1 | An Example of the Difference in Fitting Each Year Separately Rather than Fitting all the Data . . . . . | 82 |
| B-1 | Sample Scatterplot Defining Three Distinct Regions. . . . . | 85 |
| B-2 | Data from Region $R_1$ Plotted for Another Meteorological Variable, $X_2$. . . . . | 87 |
| B-3 | Decision Tree Approach Toward Defining Meteorological Classes Based on Meteorological Parameters $X_1$, $X_2$, and $X_3$ . . . . . | 89 |

## TABLES


| Number | | Page |
|---|---|---|
| 2.1 | Listing of Documented Relationships between Specified Air Contaminants and Meteorological Parameters. . . . . | 8 |
| 2.2 | Sample Seasonal Patterns of Pollutant Concentrations for Urban Areas . . . . . | 24 |
| 2.3 | Sample Diurnal Patterns of Pollutant Concentrations . . . . . | 24 |
| 3.1 | Zeldin's Point Classification Table . . . . . | 33 |
| 4.1 | Meteorological Classes Established by Weather Type for the Southern California Area . . . . . | 51 |
| 4.2 | Annual $SO_2$ Average Concentrations for the Hypothetical Receptor Site . . . . . | 68 |
| 4.3 | Tabular Layout of the $n_{km}$ $\bar{C}_{km}$ Entries and the A Priori Probability, $P_K$ for the Hypothetical Case . . . . . | 70 |
| 4.4 | Tabular Entries Showing the Product of the Top-Three-Category A Priori Probabilities, $P_K^{(2)}$, and the Average Category Concentration . . . . . | 71 |
| 4.5 | Same as Table 4.4, except over all five categories, using $P_K^{(1)}$ instead of $P_K^{(2)}$ . . . . . | 71 |
| 4.6 | Number of Annual $SO_2$ NAAQS Violations for the Hypothetical Receptor Site . . . . . | 74 |
| 4.7 | Tabular Layout for Weather Correcting the Number of Violation Days . . . . . | 74 |
| 4.8 | Normalization, by Category, of Violation Days . . . . . | 75 |

## ACKNOWLEDGMENTS

The suggestions of the EPA reviewers, particularly the project monitor, Neil Frank, have led to significant improvements in the clarity of this report.

The assistance of Leo Breiman, Yuji Horie, and others at Technology Service Corporation is greatly appreciated.

# 1.0 INTRODUCTION

Year-to-year fluctuations in meteorology can have a significant impact on measured pollutant levels. These fluctuations severely complicate the interpretation of air quality trends because of uncertainties as to whether yearly air quality changes are due to meteorology, emission changes (control strategies and/or source growth), or a combination of both. Hence, it is highly important to isolate the effects of meteorology to determine the air quality trends based on emission changes.

The purpose of this report is to provide guidelines, primarily for federal, state, and local air pollution control agencies, to assess quantitatively the effects of year-to-year meteorological variations upon air quality. The recommended procedures are designed to treat some of the major objectives of meteorological adjustment:

(1) to detect the existence and direction of a pollutant trend at a given locale;

(2) to detect the magnitude of that trend (e.g., percentage improvement/degradation); and

(3) to compare the degree of improvement in one locale to that of another.

Meteorologically adjusted ambient air quality levels provide values for each year which can be compared without the confounding effect of different meteorological conditions. Such adjusted values do not represent the trend in actual emissions as such, but the trend in emission potential. That is, even if we could measure emissions without the confounding effect of

meteorology on dispersion, there would be a confounding effect of meteorology on measured emissions. The meteorologically adjusted air quality attempts to measure the same quantity that an emission inventory attempts to measure--the normalized emissions. Since this is the quantity that can be controlled (as opposed to the weather), it is the trend that we are most interested in.

There is another complicating factor in determining meteorologically adjusted trends. If air quality has improved under all meteorological conditions, there has been clear improvement in air quality. If, however, air quality improves under certain meteorological conditions and degrades under others, there may be no clear meteorologically adjusted trend. Such a condition may occur, for example, when effective controls are placed upon one major point source on one side of an air quality monitor, and a new point source comes into being on the opposite side of the monitor. In such a case, there will be an improvement under meteorological conditions where the wind is blowing from the controlled point source toward the monitor and a degradation in air quality for meteorological conditions where the wind is blowing from the new point source toward the monitor.

The meteorological adjustment can be used to estimate the improvement in normalized emissions overall; or, in cases where there are contradictory trends, one can attempt further to isolate the components of that overall trend. We will emphasize in this report the objective of determining the overall air quality trend while discerning differences in trends among different meteorological conditions, but it is beyond the scope of this report to discuss in detail approaches which use meteorology to discern components of the trend.

## 1.1 SCOPE OF STUDY

In order to provide a foundation for performing a meteorological adjustment of air quality data, this study will evaluate the following:

(1) Meteorological Effects Upon Air Quality

Key meteorological parameters affecting specific air pollutants are listed with their known geographical region of applicability. The representativeness of such data is discussed as well as the effects of meteorology upon various air quality parameters.

(2) Approaches to Meteorological Adjustment of Air Quality Trends

While much has been written about air pollution trends, and while the effects of meteorology upon such trends have been long recognized, very little has been written concerning actual methods of weather adjustment to air quality data. A brief review of published methods is presented. The intent of this review is to indicate the key ideas in the methods suggested, with the aim of adapting the best of these ideas in the recommended methodology. Thus, the methods are not discussed in detail (except for specific technical points of interest).

(3) Recommended Procedure for Meteorological Adjustment of Air Quality Trends

We extract ideas from several of the methods reviewed to synthesize a procedure adapting and combining the best ideas in the methods surveyed. This procedure is described and exemplified in some detail in the body of this report; it is outlined in the following section of this introduction.

The intent of this report is to develop a general framework for meteorological adjustment applicable to most pollutants and most locales. It is beyond the scope of this study to develop tailored methodologies for specific averaging times, specific pollutants and/or specific locales; however, general techniques for treating a wide range of pollutants and locales are recommended. The motivation for looking at the overall problem of meteorological adjustment of air quality trends, rather than at a specific locale and a specific pollutant (as has most previous work in this area), is to provide a basis for the development of a generally acceptable approach to the problem. This basic approach is adaptable to the requirements of specific pollutants and locales.

This study was specifically intended to be an analysis of the state-of-the-art in this area, a synthesis of the best of that state-of-the-art, and a delineation of approaches generally applicable to meteorological adjustment. The study does not include empirical testing of the methodology. It is expected that experience with the approach will suggest improvements and extensions of the methodology--we do not claim to have developed a fully tested "cookbook" approach. We are, however, convinced that the recommended methodology is technically correct, practically feasible, and widely applicable. We outline this methodology in the following sub-section. The state-of-the-art and the recommended methodology are discussed in some detail in the body of this report.

Certain subjects that are relevant to the problem of meteorological adjustment could not be treated in this report. For example, we could not treat the question of detecting whether there is a statistically significant

trend following meteorological adjustment or the question of when one can claim that the differences in the meteorologically adjusted index and the original unadjusted air quality data are statistically significant.

In addition, there are some general computer methodologies which go beyond the question of meteorological adjustment of air quality trends (and, for that matter, beyond air quality problems in general) of which we have limited our discussion. For example, part of the methodology we developed involves the defintion of distinct meteorological classes relevant to distinguishing pollutant levels. We have elaborated one specific approach for the derivation of these meteorological classes which does not require specialized technical knowledge, and we have referenced other potential methods for accomplishing this classification in the literature and in terms of available software packages.

## 1.2 OUTLINE OF RECOMMENDED PROCEDURE

The procedure synthesized from previous work on meteorological adjustment and recommended in Section 4.0 has the following major characteristics:

- Meteorological variables are used to define "meteorological classes." Each class corresponds to different types of meteorological conditions which affect pollutant concentration; thus, each class will have a distinct average concentration level associated with it.
- Within each meteorological class, linear regression allows adjustment of average concentration levels to levels consistent with the "typical" meteorology for that class.

- Yearly measures of pollutant levels are adjusted to what they would be if each year had the same meteorology. The differences in yearly meteorology are reflected in the frequencies of occurrence of the meteorological classes in each year; pollution levels are adjusted to what they would be if, in each year, the frequency of occurrence of each class was the same as the average frequency of occurrence of that class over the entire period.

This brief synopsis of the recommended approach is offered to give the reader an idea of what awaits at the final destination as he navigates the long voyage to Section 4.0--and to discourage drowning and/or mutinies along the way.

# 2.0 METEOROLOGICAL EFFECTS UPON AIR POLLUTION

## 2.1 METEOROLOGICAL VARIABLES AFFECTING AIR POLLUTION

Meteorological adjustment of air quality data is predicated upon significant relationships between pollutant and one or more meteorological parameters. The selection of appropriate parameters depends upon the availability of meteorological data and the air quality index of interest. Many experienced air pollution agency personnel have intuitive knowledge of key meteorological parameters. For example, light wind conditions usually result in increases in pollutant levels, while gusty winds generally are associated with excellent ventilation and reduced pollution. In some areas of the country, increasing surface temperatures foretell of increasing photochemical pollution. It is this type of "well-grounded" intuition which can be of great value in selecting appropriate meteorological variables. The following section presents a comprehensive listing of known relationships between meteorological and air quality data. This is intended to be a guide toward selection of appropriate meteorological parameters. Beyond this, classical statistical techniques are necessary to carry forward the selection process.

### 2.1.1 Key Meteorological Variables by Pollutant

This section is intended to provide an overview of some of the relationships between air pollutants and specific meteorological parameters which have been determined in prior research. While this compilation does not list every parameter ever tested, it does present a comprehensive guide for use in selecting the key meteorological parameters. Table 2.1 denotes the key meteorological parameters by types of pollutants. Items which are preceded by a ⊕ or ⊖ indicate those parameters which have

Table 2.1 Listing of Documented Relationships between Specified Air Contaminants and Meteorological Parameters. (Positive correlations are indicated by "+", while negative correlations are denoted by "-". Significant correlations are represented by ⊕ or ⊖ preceding the meteorological parameter. Non-linear and insignificant relationships are also noted.)

| | METEOROLOGICAL PARAMETER | RELATIONSHIP | GEOGRAPHICAL AREA | REFERENCE |
|---|---|---|---|---|
| | POLLUTANT: OXIDANT | | | |
| ⊕ | 1. Solar Radiation | + | New York & New Jersey | 3 |
| | | | Los Angeles | |
| ⊖ | 2. Wind Speed | - | New York & New Jersey | 16,23 |
| | | - | St. Louis | 23 |
| | | - | Los Angeles | 16 |
| ⊕ | 3. Maximum Temperature | + | New York | 4 |
| | | + | St. Louis | 23 |
| | | + | Los Angeles | 16,25 |
| | | + | San Francisco | 1 |
| | 4. Mixing Height | Insig. | New York | 4 |
| ⊕ | 5. Wind Direction | | New York | 4 |
| | 6. Days Since Precipitation | + | St. Louis | 23 |
| | 7. Days Since Frontal Passage | + | St. Louis | 23 |
| | 8. Change in 500-mb Height | + | St. Louis | 23 |
| | | + | Los Angeles | 12,14,24 |
| | 9. 500-mb Height | + | St. Louis | 23,25 |

Table 2.1 (Cont'd)

| METEOROLOGICAL PARAMETER | RELATIONSHIP | GEOGRAPHICAL AREA | REFERENCE |
|---|---|---|---|
| POLLUTANT: OXIDANT (Continued) | | | |
| 10. $\Delta T_{(30m-5m)}$ | Insig. | St. Louis | 16 |
| 11. Relative Humidity | Insig. | Los Angeles | 16 |
| 12. Stability Class | Insig. | Los Angeles | 15 |
| ⊖ 13. Inversion Base | - | San Francisco | 1 |
| | - | Los Angeles | 25,44 |
| 14. Inversion Magnitude | + | Los Angeles | 25,44 |
| ⊕ 15. 850 mb Temperature | + | Los Angeles | 5,12,26,44 |
| 16. 950 mb Temperature | + | Los Angeles | 44 |
| 17. Pressure Gradient | Non-linear | Los Angeles | 44 |
| 18. Mixing Height x Wind Speed | - | Los Angeles | 15 |
| 19. Temperature at 2500 ft. | + | Los Angeles | 14 |
| POLLUTANT: CO | | | |
| 1. Inversion Base | - | Los Angeles | 35 |
| 2. Mixing Height | - | Los Angeles | 35 |
| 3. Wind Speed | - | Los Angeles | 35 |

Table 2.1 (Cont'd)

| METEOROLOGICAL PARAMETER | RELATIONSHIP | GEOGRAPHICAL AREA | REFERENCE |
|---|---|---|---|
| POLLUTANT: $NO_2$ | | | |
| 1. Mixing Height | - | Los Angeles | 37 |
| 2. Maximum Temperature | + | Los Angeles | 39 |
| 3. Minimum Relative Humidity | Insig. | Los Angeles | 39 |
| 4. Wind Speed | - (winter) | Los Angeles | 39 |
| | Insig. (summer) | | 39 |
| 5. Solar Radiation | Insig. | Los Angeles | 39 |
| 6. Pressure Gradient | Insig. | Los Angeles | 39 |
| POLLUTANT: TSP | | | |
| 1. Wind Speed | - | Los Angeles | 16 |
| | - | New York | 16 |
| 2. Relative Humidity | Insig. | New York | 16 |
| 3. Temperature | Insig. | New York | 16 |
| $\Delta Temp_{(Max-Min)}$ | + | Cleveland | 33 |
| Minimum Temperature | + | Cleveland | 33 |
| 4. Precipitation | Insig. | New York | 16 |
| | - | Los Angeles | 41 |

Table 2.1 (Cont'd)

| METEOROLOGICAL PARAMETER | RELATIONSHIP | GEOGRAPHICAL AREA | REFERENCE |
|---|---|---|---|
| POLLUTANT: TSP | | | |
| 5. Pressure | + | Cleveland | 33 |
| 6. Precipitable Water | - | Cleveland | 33 |
| 7. Degree Days | + | Boston | 19 |
| 8. 850-mb Temperature | + | Los Angeles | 11 |
| POLLUTANT: $SO_2$ | | | |
| 1 Degree Days | + | Boston | 19 |
| 2. Wind Speed | - | Boston | 19 |
| | - | Los Angeles | 41 |
| | - | New York | 40 |
| | - | Netherlands | 43 |
| 3. Temperature | Non-linear | New York | 40 |
| | + | Netherlands | 43 |
| 4. Mixing Height | - | Netherlands | 43 |

Table 2.1 (Cont'd)

| METEOROLOGICAL PARAMETER | RELATIONSHIP | GEOGRAPHICAL AREA | REFERENCE |
|---|---|---|---|
| POLLUTANT: $NO_x$ | | | |
| 1. Wind Speed | | Los Angeles | 22 |
| 2. Inversion Base | | Los Angeles | 22 |
| 3. Wind Direction | | Los Angeles | 22 |
| POLLUTANT: $SO_4^=$ | | | |
| 1. Inversion Base | Non-linear | Los Angeles | 45, 47 |
| 2. Inversion Magnitude | Non-linear | Los Angeles | 45 |
| 3. 7000-mb Dewpoint | | Los Angeles | 47 |

been shown to exhibit similar relationships to a specific contaminant over several geographical locations or have had significant relationships in one area. Further, this section discusses, by pollutant, the known logical cause-effect relationships.

Photochemical Oxidants (Ozone)

Meteorological effects upon the formation and accumulation of oxidants are well documented, primarily because this particular pollutant has been the subject of intense investigation. Since photochemical reactions upon primary pollutants ($NO_x$ and HC) occur in the presence of ultraviolet radiation (UV), oxidant concentrations and available solar radiation have been found to be positively correlated [3,16,23]. Despite the good relationship, intense solar radiation must be coupled with other meteorological conditions to cause high oxidant conditions. For example, in the Los Angeles area, a hot, desert wind condition known as "Santa Ana" winds, is associated with abundant solar radiation, but the overall meteorological condition carries the pollutants out to sea.

A related parameter, surface temperature, is a function of solar radiation and air-mass characteristics. Accordingly, temperature also correlates positively with oxidants [1,4,16,23,25]. Generally, as the surface temperature increases, oxidant levels increase. However, there is usually some limiting temperature above which conditions become atmospherically unstable, thus allowing pollutants to escape vertically.

Low wind speeds correlate with elevated oxidant levels [3,4,16,23] because stagnant early-morning conditions represent a greater potential for accumulating primary pollutants. While this condition is essential for resulting in high oxidant concentrations within major source areas, it does not necessarily hold true for outlying receptor areas. In fact, in analyzing data in a distant receptor location, strong, persistent winds blowing from the source area may be a prerequisite for resultant high oxidant concentrations. As illustrated in Figure 2.1, data from San Bernardino (located approximately 60 miles east of Los Angeles) exhibit a strong relationship between increased wind velocity and increased oxidant levels.

Limited mixing conditions as a result of low inversion base heights have been shown to be of considerable importance in causing elevated oxidant levels [1,25,44]. (For a discussion of "inversion" and "mixing height," see Section 2.1.3). However, surface-based inversions tend to break under surface heating more readily than persistent, elevated inversions. Once the inversion is broken, a much greater potential for mixing is achieved. It is therefore advisable to treat surface inversion conditions as a separate condition when relating mixing height to oxidant concentrations.

Temperature and pressure conditions aloft (i.e., 850-mb; 500-mb levels) have been shown to correlate with oxidant levels [5,12,14,23,24,25,26,44]. Conditions that relate to a synoptic weather pattern associated with a warm dome of high pressure aloft, usually result in weak ground-level circulation under a very stable atmosphere. In this regard, warm temperatures aloft (associated with much-above-average contour heights) at the various atmospheric levels are conducive to elevated oxidant levels.

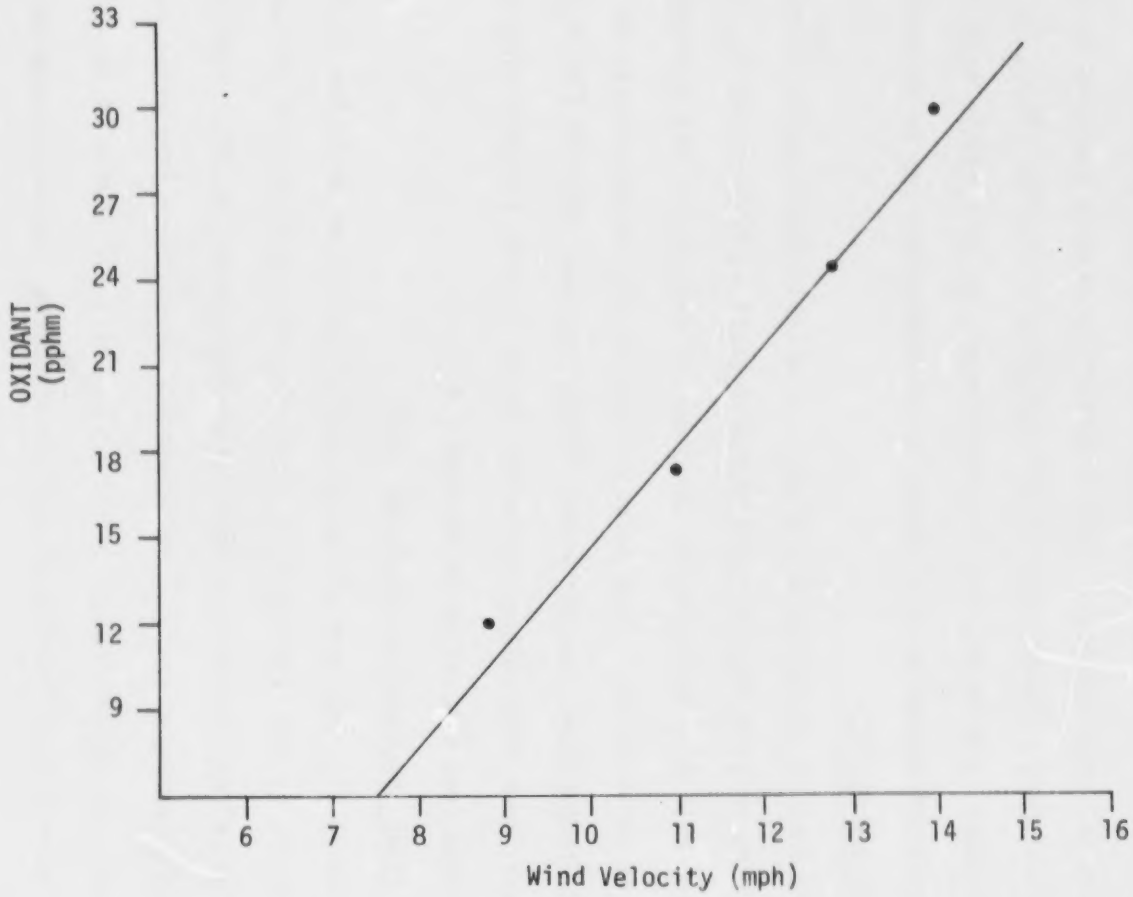


Figure 2.1 San Bernardino Smog Season: Maximum Oxidant vs. Velocity at Time of Occurrence. Dots represent the means of grouped classes. (Courtesy of the South Coast Air Quality Management District.)

Carbon Monoxide

Unlike photochemical oxidants which require examination of meteorological conditions conducive to primary pollutant accumulation, photochemical reactions, and transport, carbon monoxide build-up generally occurs only under extremely stable, surface inversion conditions. Inversion base height and wind speed have been shown to correlate with CO concentration levels [35]. These parameters may help define a type of overall weather pattern resulting in increased CO concentrations. Such a pattern is usually typified by clear, windless nights causing significant radiational cooling. Thus, by examining temperature, wind speed, and inversion depth, it is possible that the type of condition conducive to elevated CO concentrations can be determined.

Sulfur Dioxide

Because $SO_2$ emissions are almost exclusively point-source oriented, the governing relationships between meteorological parameters and $SO_2$ concentration will vary according to the specific features of the source(s), such as stack height. Some success relating $SO_2$ concentrations to wind speed, temperature, and/or mixing height has been obtained [19,40,41,43]; however, some difficulties can arise due to extreme variabilities in $SO_2$ emissions (see discussion in Section 2.2.1).

Total Suspended Particulates (TSP)

Basically, the type of meteorological condition conducive to increased TSP levels is the same stagnant classification necessary for the accumulation of primary pollutants, namely low wind speeds [16,19]. However, the dependency upon wind velocity is a function of distance from major source regions. As in the case of oxidant transport, TSP levels in non-urban areas may increase as a function of the persistence of source-to-receptor winds.

As a further complication, any dry, windy area may experience the greatest particulate problem under extremely strong wind conditions when natural particles are carried aloft (i.e., duststorm conditions).

For urban areas, stable, anticyclonic conditions are prerequisites for high TSP levels. Some success in correlating to TSP has been shown with temperature [16,33], pressure [15] and 850-mb temperatures [11].

### 2.1.2 Availability of Meteorological Data

Perhaps the single most limiting factor in performing successful air quality data adjustment is the availability of meteorological data. Certain areas of the country, primarily the larger urban centers, are fortunate to have extensive sources of meteorological data. Other areas more removed from the urban complexes are subject to limited meteorological data.

The following list will serve as a guide to potential sources of meteorological data:

Upper Air Data

National Climatic Center (catalogue listing #TDF 56)

Research Libraries (e.g., Daily Weather Map Series)

Surface Data

National Climatic Center (catalogue listing #TD1440)

Pollution Control Agencies

Flood Control Agencies

U.S. Forest Service stations

Schools/Universities

### 2.1.3 Representativeness of Meteorological Data

The success of any methodology for producing weather-adjusted air quality trends depends upon the representativeness of available meteorological data. "Representativeness", in this context, implies those meteorological data which can adequately define meteorological conditions at the air monitoring location. Ideally, both meteorological and air quality data, measured at the same location, provide the best combination of conditions. Usually, however, such data are not measured at common locations (or at least a variety of meteorological parameters are not measured at the same location). It then becomes a problem of locating sources of meteorological data and the subsequent selection of appropriate parameters.

This section will describe some of the conditions necessary to estimate the representativeness of meteorological data from sources which are not co-located with available air quality data. Although the treatment of data "representativeness" is of proportions to warrant a separate study (currently in progress by the EPA), an overview is presented here in order to highlight pertinent considerations supporting meteorological adjustment techniques. Meteorological parameters will be considered with respect to three conditions: (1) horizontal distance, (2) topographical effects, and (3) weather-type effects.

1. Surface temperature

In general, surface temperature is an excellent meteorological parameter to utilize, partly due to the availability of data and partly due to known correlations with specific contaminants. In flat terrain, isolated from land/sea breeze effects, surface temperature can be representative for large distances (> 100 km). In irregular terrain, caution must be taken to ensure

the representativeness of the temperature data; however, it is possible that data from an elevated location may serve as an indicator of warming or cooling aloft, which in turn can affect the stability of the air mass.

Locations near large bodies of water are not usually representative of locations 10-25 km inland when local sea breezes prevail. It is recommended that sources of temperature data be used which represent similar micro-climatology.

2. Surface winds

Perhaps the most consistently usable meteorological parameter is surface winds. Relationships between low wind speeds and pollutant concentration increases are well documented. Surface winds can also be used to define source-receptor transport patterns. Consequently, wind data from distant locations can be useful, if pollutant transport is a primary mechanism for affecting a particular monitoring site.

Extreme variability in wind conditions can occur in areas with elevated terrain and, to a lesser extent, in micro-meteorological regimes (i.e., land/sea breezes). It is not advisable, therefore, to use a single source of wind information. Rather, an area composite of available stations should be used.

3. Upper air data

The major advantage of using upper atmospheric temperature/humidity/and wind data is the tendency toward homogeneity of conditions at or above 850 mb (~1500 m). It is sufficient to say that such conditions are probably representative of areas within a 100-km radius. Because the horizontal variations in meteorological conditions tend to decrease with altitude, such parameters may be most useful in mountainous terrain, where surface measurements may reflect extremely localized conditions.

## 4. Inversion and mixing height data

Unlike standard atmospheric conditions, where temperature decreases with height, a temperature inversion is defined as a layer of air in which temperature increases with height. Thus the lifting of an air parcel from the surface, undergoing adiabatic cooling, is inhibited from rising above the inversion due to the changes in buoyant properties between the parcel and its surrounding environment. This is especially important in air pollution considerations because of the vertical dilution limitations caused by the inversion.

Another term reflecting limitations to vertical dilution is "mixing height." In this report, "mixing height" is defined as that level at which an air parcel with a given surface temperature is lifted and cooled at the adiabatic lapse rate of 9.8°C $km^{-1}$ until it achieves the same temperature as the surrounding ambient air with a more stable lapse rate. The existence of a definite mixing height does not imply the existence of an inversion; however, the existence of a low-level inversion does identify limited mixing.

When inversion/mixing height data are available, definite implications to pollutant concentrations can be made. While vertical air-mass properties may be representative within 100 km, the effects of mixing are governed by surface heating (i.e., as the surface temperature increases, the level at which equilibrium occurs between the lifted air parcel and the ambient air also increases) and general atmospheric stability. As a result, the cautions mentioned for the widespread geographical use of temperature parameters also apply for inversion/mixing height data.

5. Weather types

No one single meteorological parameter can adequately describe the prevailing weather conditions at the time of a particular air quality measurement. To get a true picture of the cause-effect phenomena of meteorology upon air quality, several independent or interrelated variables should be gathered to obtain the most representative meteorological assessment at a given air quality monitoring location. While individual parameters may not be representative due to terrain or localized influences, their use with other parameters may be significant. For example, a certain mixing height may not indicate a definite relationship to a pollutant concentration; however, the mixing height coupled with temperature, relative humidity, and wind speed may perfectly describe a weather condition which results in a consistent pollutant distribution and concentration. It suffices to say that combinations of meteorological variables tend to become more representative of dominating weather types than the individual parameters, provided that logical meteorological processes are considered.

No specific list describing the representativeness of available meteorological data can be compiled for every possible location across the country. For every rule given, there will always be an exception. The general "common sense" guidelines presented here are intended to stimulate the reader toward a goal of thoughtful, rather than haphazard, selection of available meteorological data. An empirical method is presented in Section 4.0 to assist in the selection process.

## 2.2 VARIATIONS IN AIR QUALITY DATA DUE TO METEOROLOGY

The relationship of specific pollutants to various meteorological parameters, cited in Table 2.1, suggest that variations in pollutant concentrations, especially on a daily basis, are primarily due to meteorology. Except in the case of specific point sources, where daily changes in emissions can affect air quality in a substantial manner, the general uniformity in daily emissions over most urban areas dictates that short-term changes of measured concentrations are caused by meteorological fluctuations. To quantify the amount of variation attributed to meteorology is a difficult and complex problem. Certainly if techniques using meteorological parameters could explain all of the daily pollutant variance, then one would have achieved a perfect forecast mechanism. Nothing approaching this level of achievement has been accomplished to date. This does not imply that such near-perfect relationships do not exist; but rather the implication is that the explicit combination of meteorological parameters necessary to determine the variation is too complex to identify and measure exactly.

The longer the period of analysis, the greater the potential for pollutant variances to be complicated by both meteorological and emission factors. For example, in one study [39], the statistical variance for year-to-year nationwide $NO_2$ values during 1970-1975 was found to be ±20% for the annual 1-hour maximum, ±13% for the 99th percentile of hourly concentrations, and ±11% for the annual mean. No attempt, however, was made to distinguish that part of the total variance due to meteorology, illustrating the complexity of the problem.

While the quantitative assessment of meteorological variations affecting a variety of pollutants can perhaps best be dealt with in a lengthy investigation of historical variations of specific meteorological parameters (beyond

the scope of this report), the following subsections will explain some of the known meteorological factors affecting air pollutant concentrations as well as the subsequent effects inherent in the type of air quality parameter(s) chosen for analysis.

2.2.1 Variation in Meteorological Effect Among Pollutants

The effect of meteorology on pollutant concentrations may differ from one pollutant to another. For example, higher wind speeds tend to result in lower concentrations for primary pollutants such as CO. But this is not necessarily true for oxidants or $SO_2$ which may be affected to a greater degree by large point sources and transport phenomena. In addition, seasonal effects can account for significant variability of pollution concentration. For example, in the winter, $SO_2$ concentrations tend to increase, but oxidant concentrations decrease.

Table 2.2 presents sample seasonal patterns of three primary pollutants resulting from fuel combustion, $SO_2$, CO, and NO, and that of the secondary pollutant, oxidant. In Table 2.3 sample diurnal patterns of the same pollutants as those in Table 2.2 are presented. From these two tables, one can deduce a role of meteorology in pollutant concentration levels.

The three fuel combustion pollutants, $SO_2$, CO, and NO, exhibit higher concentrations in winter and lower concentrations in summer. This is a function not only of meteorological conditions conducive to pollutant build-up, but also of emissions. Colder winter temperatures bring greater fuel-burning requirements, which in turn lead to greater seasonal differences in pollutant concentrations. On the other hand, the diurnal pattern of these fuel combustion pollutants may be explained more by meteorology than by emissions. The concentrations of the three pollutants generally increase in

Table 2.2 Sample Seasonal Patterns of Pollutant Concentrations For Urban Areas

| Pollutant | Spring (M,A,M) | Summer (J,J,A) | Fall (S,O,N) | Winter (D,J,F) |
|---|---|---|---|---|
| $SO_2$ | Medium | Low | Medium | High |
| CO | Medium | Low | Medium | High |
| NO | Medium | Low | Medium | High |
| $O_x$ | Medium | High | Medium | Low |

Table 2.3 Sample Diurnal Patterns of Pollutant Concentrations

| Pollutant | Morning (6-12) | Afternoon (12-18) | Evening (18-24) | Night (0-6) |
|---|---|---|---|---|
| $SO_2$ | High | Low | Medium | Medium |
| CO * | High | Medium | High | Medium |
| NO | High | Low | Medium | Medium |
| $O_x$ | Medium | High | Low | Low |

* CO typically (but not at all locales) has three peaks: around 0700; around 1700; and, with surprising regularity, around 2400.

the morning and decrease in the afternoon, despite the fact that the emission levels remain high in that period. The probable cause of this difference in morning and afternoon concentrations will be found in meteorology. Under stable, limited mixing, early-morning meteorology, concentrations will increase. In the afternoon, solar radiation and wind speeds increase the atmospheric ventilation, thus resulting in dilution of pollutant concentrations.

The secondary pollutant, oxidant, is formed by complex photochemical reactions of precursor pollutants, hydrocarbons and oxides of nitrogen. The oxidant concentration level is higher in summer than in winter (Table 2.2), and is higher in the afternoon than in the morning (Table 2.3). The high oxidant concentrations during the summer season and afternoon period are primarily due to the unfavorable meteorology (high temperature and strong insolation) during those periods rather than due to the slightly increased hydrocarbon emissions (i.e., increase in evaporative emissions).

Smelters and power plants are often responsible for much of the $SO_2$ emissions. Peak $SO_2$ concentrations occur when the wind blows from these point sources to a recepter site. Therefore, $SO_2$ concentrations originating from point sources are more sensitive to wind direction than wind speed. To further complicate matters, breakdowns of modern, efficient pollution control equipment can result in substantially higher concentrations at a monitoring location under a variety of meteorological conditions. Thus in certain areas of the country, such as Los Angeles, current $SO_2$ concentrations may be more of a function of industrial control-equipment-breakdown status than of any specific set or sets of meteorological conditions.

Consideration must be given to the elevation of point-source emissions. Ground-level emissions, such as CO from vehicular traffic, would be affected by meteorological conditions which caused strong low-level inversions with very light wind speeds. Stack emissions could easily penetrate those types of inversion conditions. Strong off-the-surface inversions might trap stack emissions, but would not be conducive to high concentrations of ground-based emissions.

The type of pollutant examined will affect the horizontal difference and geographical complexity of the meteorology to be considered. The concentration of primary pollutants is determined by local emissions and local meteorological conditions. Therefore, meteorological adjustment of air quality trends for the primary pollutants should be made by using the local meteorological data collected as near as possible to the air monitoring station. On the other hand, the concentration level of secondary pollutants is determined by regional emissions and meso-scale or synoptic scale meteorological conditions. In the long-range transport of secondary pollutants, such as oxidant, it may be necessary to examine the meteorological conditions nearer the source area than the receptor area. In some analyses, this may require investigation of meteorological data 100 km or more from the monitoring location.

### 2.2.2 Variation in Meteorological Effect Among Air Quality Parameters

Currently used air quality parameters can be categorized into three types: mean concentration at various averaging times; extreme concentrations such as daily maximum, annual maximum or annual second maximum; and frequency of

hourly or daily maximum hourly concentrations exceeding air quality standards. Meteorology may affect these air quality parameters differently.

The level of a maximum concentration is largely dominated by a particular meteorological condition, rather than an emission level. Variation in annual 1-hour maximum concentrations reflects the variability in annual "worst-case" meteorology. On the other hand, the level of an average concentration is a more stable parameter: the longer the averaging period, the more stable the parameter. Anomalous meteorological years tend to be smoothed out in annual average pollutant statistics, although significant seasonal anomalies can seriously affect the results. Regardless of the parameter selected, it is difficult to extract direct emission-related air quality trends without an examination of the underlying meteorological conditions.

Similarly, the number of violations of a standard, or the frequency of concentrations exceeding a given threshold level can be affected in much the same manner as a maximum concentration or an average concentration, depending upon the level of the standard or that of the threshold chosen. When the threshold level chosen is very high, frequency of concentrations exceeding the threshold may vary greatly from year to year, similar to the annual maximum concentration. When the threshold level is low, the frequency of concentrations exceeding the threshold may vary as that of an average concentration.

Whichever air quality parameter is used, a specific set of logical meteorological parameters must be specified. It is rather unlikely that an identified meteorological parameter used in analyzing an air quality parameter (such as annual average concentrations) could be used to explain another air quality parameter. For example, an abnormally large number of winter surface

inversions might explain an increase in annual CO averages, but could not, of itself, explain an unusually high single CO hourly average. Specific meteorological parameters must be identified to explain the individual case.

## 3.0 APPROACHES TO METEOROLOGICAL ADJUSTMENT OF AIR QUALITY TRENDS

The adjustment of observed air quality for meteorological variability to reflect true trends requires three basic elements:

(1) Selecting an air quality index: What annual measurement of air quality do we wish to adjust? Candidates include (a) values corresponding to the standards (e.g., the second-highest eight-hour average CO value in the year); (b) averages of values related to the standards (e.g., the average of daily maximum eight-hour-average CO over the year); (c) the averages of values related to the standards over a portion of the days of the year (e.g. only the high-pollution season);(d) the number of days or hours of the year in which the standard is exceeded; and (e) variations on the above themes.

(2) Relating meteorological variables to air quality: If air quality measurements are to be adjusted for meteorology, the effect of available meteorological variables on air quality must be expressed at least qualitatively, but preferably quantitatively.

(3) Adjusting measured air quality: The relationship between air quality and meteorological variables must be used to adjust observed values. Preferably, the adjusted values should represent annual pollution measurements which would have occurred if meteorology were the same each year.

These three components of a meteorological adjustment procedure are closely interrelated; the best choice of one component may depend on the choices we make of the other components. In this section, we discuss some choices suggested in the literature, indicate their relationship to one another, and outline their strengths and weaknesses. In Section 4.0, we recommend an approach which is an extension of the best ideas in the literature.

The remainder of this section is organized by the way in which the proposed methodology relates meteorology to air quality:

(a) Postulation from basic principles;

(b) Regression techniques;

(c) Classification techniques; or

(d) Meteorologically invariant statistics.

## 3.1 POSTULATION FROM BASIC PRINCIPLES

Section 2.0 discusses qualitatively typical relationships of individual meteorological variables to the concentration of various pollutants. Using these and similar ideas, one might combine meteorological variables to develop an "index" indicating the likelihood of high pollution levels in a given day or year.

For example, an Environmental Protection Agency (EPA) report [42] used a method of meteorological adjustment of oxidant trends in the Los Angeles area which compared annual ventilation criteria and yearly hydrocarbon emissions to the annual average of daily one-hour oxidant maxima (a composite of three key monitoring locations). Ventilation data represented the number of days per

year with specified limited mixing and low wind speeds (Los Angeles County APCD "Rule 57" days). A day meets the "Rule 57" criteria if the following all occur on that day:

(1) Morning inversion < 1500 feet MSL;

(2) Afternoon mixing height $\leq$ 3500 feet MSL; and

(3) 6 A.M. - 12 noon average wind speed $\leq$ 5 mph.

One potential difficulty in the annual agglomeration of Rule 57 days against annual oxidant data is that poor ventilation in the off-season winter months does not relate well to high concentrations of oxidant. As a result, seasonal biases can occur by achieving an unusually high number of winter Rule 57 days or an unusually low number of summer Rule 57 days. This problem is partly solved by treating the "smog season" alone rather than the full year.

Horie and Trijonis [20] used a conceptually similar approach to estimate the effect of meteorology on annual values of oxidant and $NO_2$ in Los Angeles. If a parameter value in a given year was sufficiently above its long-term average, it was assigned a "+1" value; if sufficiently below, a "-1" value. A "score" was calculated for each year to indicate which years had an above- or below-average propensity for high pollutant concentrations. The results were used qualitatively to comment on the likely underlying pollution trends.

Another index was developed by Davidson [15] using the product of mixing height and wind speed in Los Angeles. This combination reflected atmospheric conditions conducive to photochemical pollution. Thus, the lower the index value, the greater the seasonal potential for increased oxidant concentrations. Ozone trends in the Los Angeles area were normalized based on the seasonal index value. Results showed that adjusted

seasonal daily one-hour maxima were improving slightly less rapidly than actual trend data due to the more favorable meteorological conditions in the early 1970's.

An index for $SO_2$ near Rotterdam was used by van Dop and Kruizinga [43] based upon a measure of the strength of the inversion ($\Delta\theta$), the maximum mixing height (L), and the average daytime surface windspeed (U). The form was

$$A = C_1 \Delta\theta + C_2/LU \quad ,$$

where $C_1$ and $C_2$ were chosen so that the two terms have equal values for average conditions. The daily value of A was averaged over each year for use in normalizing that year.

Zeldin [44,46] developed a method for meteorological adjustment using weighted class-intervals for specific meteorological parameters. Based on known cause-effect relationships between oxidant and meteorological parameters of inversion strength, temperatures aloft, inversion height, and surface pressure gradients, a stratified system was devised which numerically evaluated the meteorological potential for each day. Table 3.1 illustrates the "scoring" structure. A high score indicated a high pollution potential.

Zeldin averaged daily meteorological index values for the smog season months of June through September. Resultant values were therefore a numerical evaluation of the severity of the entire smog season. Statistical comparisons were then made to oxidant data for various cities in the Los Angeles Basin. In one study [44], correlations were performed using the number of hours exceeding the California Air Resources Board (ARB) stage-one episode level of 0.20 ppm.

| Point Score | (°C) Inv. Strength | (°C) 950-mb Temp. | (feet) Inversion | (millibars) Pressure Gradient |
|---|---|---|---|---|
| 0 | ≤ 5.0 | ≤ 5.0 | 5,000+ | ≤ -10.0<br>≤ +20.0 |
| 1 | 5.1 - 7.0 | 5.1 - 8.0 | 4,001 - 5,000 | - 9.0 to - 9.9<br>+16.0 to +19.9 |
| 2 | 7.1 - 9.0 | 8.1 - 11.0 | 3,001 - 4,000 | - 8.0 to - 8.9<br>+12.0 to +15.9 |
| 3 | 9.1 - 11.0 | 11.1 - 14.0 | 2,501 - 3,000 | - 7.0 to - 7.9<br>+12.0 to +11.9 |
| 4 | 11.1 - 13.0 | 14.1 - 17.0 | 2,001 - 2,500<br>Surface | - 6.0 to - 6.9<br>+ 8.0 to + 9.9 |
| 5 | 13.1 - 15.0 | 17.1 - 20.0 | 1,501 - 2,000 | - 5.0 to - 5.9<br>+ 6.0 to + 7.9 |
| 6 | 15.1 - 17.0 | 20.1 - 24.0 | 1,001 - 1,500 | - 4.0 to - 4.9<br>+ 4.0 to + 5.9 |
| 7 | 17.1 - 19.0 | 24.1 - 28.0 | 701 - 1,000 | + 2.0 to + 3.9 |
| 8 | 19.1 - 21.0 | 28.1 - 32.0 | 501 - 700 | - 3.0 to - 3.9<br>+ 1.0 to + 1.9 |
| 9 | 21.1 - 23.0 | 32.1 - 36.0 | 301 - 500 | - 2.0 to - 2.9<br>0.0 to + 0.9 |
| 10 | ≥ 23.1 | ≥ 36.1 | 150 - 300 | - 1.9 to - 0.1 |

Table 3.1 Zeldin's Point Classification Table. (Each of the four meteorological variables yielded a score. The daily score was the sum of the scores for each variable.)

In a second study [46], similar evaluations were made using the averages of the daily maximum hourly average. Results for common locations with both air quality parameters exhibited similar trends.

Using the postulation approach, some of the indices evolve in rather arbitrary units which cannot be related directly to pollutant concentrations. To achieve a more desirable result, most of the referenced authors used a single-variable linear regression of the annual pollution values against annual values (or averages) of the meteorological index in order to translate arbitrary meteorological index values into pollution concentration units (Fig. 3.1). Suppose the resulting best-fit equation is

$$y = ax + b \quad , \tag{3-1}$$

where y is the annual pollution measure and x the annual meteorological index. Then this equation is intended to give the value of y which would be "expected" to result in a year when meteorology yielded a value x. It is assumed that the differences between the expected value and observed value in a given year (the "residuals") are primarily the result of emission trends. Zeldin [44,46] used this concept directly and plotted the residuals to elucidate the true trend.

Most of the other references cited in this section took the more preferable approach of attempting to indicate what the pollution level would have been in each year if the meteorology index took its average value each year. The references are uniform in a lack of specificity as to how this adjustment of measured values was accomplished. It appears that the procedure used is as follows:* Let $\bar{x}$ be the overall mean value of

*Verified by conversation for two references [15,41].

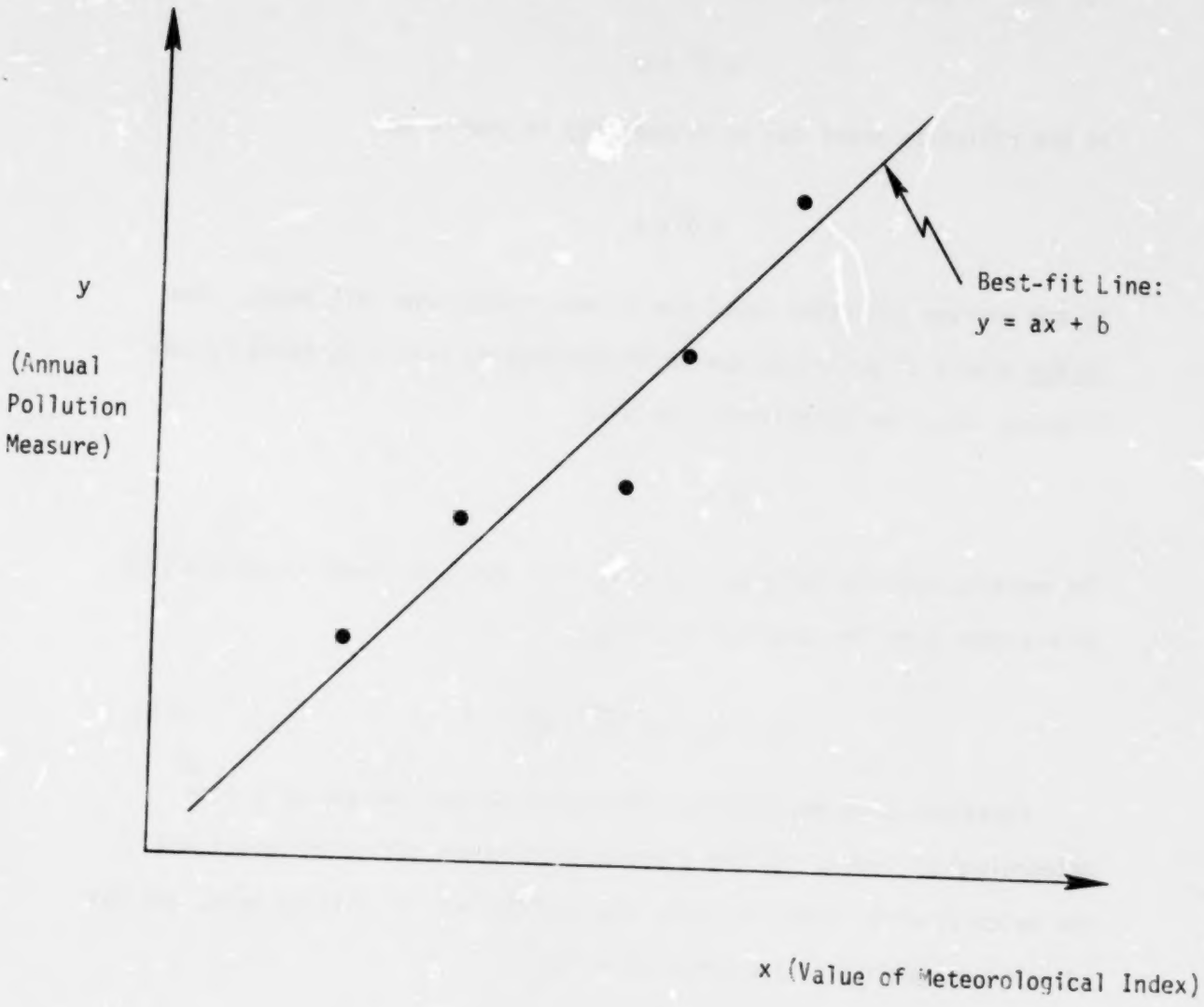


Figure 3.1 Annual Value of Pollution Measure Versus Annual Values of Meteorological Index. (Five years are assumed in this figure.)

the meteorological index (over all years), and $\overline{x}^k$ the average value for the $k^{th}$ year. Assuming Eq. (3-1) holds,

$$a\,\overline{x}^k + b$$

is the pollution level due to meteorology in year k and

$$a\,\overline{x} + b$$

is the average pollution level due to meteorology over all years. The excess amount of pollution due to meteorology in year k is found by subtracting these two quantities, yielding

$$a\,(\overline{x}^k - \overline{x}) \quad .$$

The meteorologically adjusted value $\hat{y}_k$ for year k is found by subtracting this excess from the observed value $y_k$:

$$\hat{y}_k = y_k - a\,(\overline{x}^k - \overline{x}) \quad . \qquad (3\text{-}2)$$

Equation (3-2) has built-in insurance against the use of a poor meteorological index. If the correlation between pollution levels and the meteorological index is zero, the coefficient "a" will be zero, and the adjusted value will be the observed value.*

---

*Equation (3-2) can be interpreted as the residual error in the regression for year k,

$$y_k - a\,\overline{x}^k - b \quad ,$$

added to the overall average of y,

$$\overline{y} = a\,\overline{x} + b \quad .$$

Thus, except for a change of origin, a plot of the residuals will look exactly like a plot of values adjusted by Eq. (3-2).

The major technical weakness of this approach is that the slope of the regression line (the constant "a") is affected by the underlying emission trend, as well as by meteorology. To see that this is true, one need only examine the extreme case where only two years are considered. Since the regression line in this case will go through both points, the residuals will be zero. Thus, this method will always indicate no trend in the meteorologically adjusted values, irrespective of their value, for two years. This problem is discussed in more general terms in the following subsection.

## 3.2 THE REGRESSION APPROACH

In the previous subsection, we described meteorological indexes which were postulated from basic principles. The underlying objective of this subsection is the derivation and use of such indexes by objective methods.

### Meteorological Adjustment for Two Time Periods

A study of $SO_2$ trends in Oslo, Norway, is typical of this approach [17]. A study was undertaken to meteorologically adjust $SO_2$ air quality to compare the periods 1959-1963 and 1969-1973. The meteorological conditions during the former period were considerably different from those during the later period; hence one could not expect a change in air quality to be directly related to a change in emission potential.

Data from the earlier period (1959-1963) were used to do a linear regression analysis for daily values. It was discovered that two variables dominated the estimate of $SO_2$ concentration: (1) a temperature difference between a low-altitude and high-altitude measuring station, and (2) the temperature at the lower station. For example, a typical regression equation for one station was

$$q_{SO_2} = 61.5\ (T_2-T_1) - 11.6T_1 + 472 \quad , \qquad (3\text{-}3)^*$$

where

$q_{SO_2}$ = daily mean value of $SO_2$ concentration in $\mu g/m^3$ at the particular station

$T_2$ = temperature at higher station at 7 P.M.

$T_1$ = temperature at lower station at 7 P.M.

The temperature-difference term is related to the strength of the inversion and hence the dispersion, while the temperature term is related to the variation in the emission of $SO_2$ due to space heating. Since the daily temperature data for the later time period is known, the daily mean $SO_2$ expected for the meteorological conditions during that time period can be estimated by Eq. (3-3). This was done for each day on which data was available in the later time period; the predicted values were consistently higher than observed values. Thus, it was easy to conclude that emission potential was reduced.

A quantitative statement was made in the report that the $SO_2$ pollution was reduced 50%to 60%. According to a conversation with one of the authors of the report, this latter statement was derived by looking at the ratio of the coefficient on the temperature-difference term in the early time period

---

*This equation explained the observed values of $SO_2$ concentration with a multiple correlation coefficient of .80; that is, the correlation between values predicted by this equation and observed values for the period was 0.80.

to the coefficient of the temperature-difference term in a similarly derived equation for the later time period. The intuitive justification for such a statement is that the coefficient measures the degree to which a given temperature inversion will be translated into $SO_2$ concentrations. Thus a 50% or 60% reduction in that coefficient might be thought of as a meteorologically adjusted measure of the trend in air quality. This approach to quantifying the reduction is not easily generalized or of clear mathematical validity.

Meisel suggested a more general approach to comparing the air quality in two time periods [28]. He suggested that equations relating daily air quality to daily meteorology be derived for the two time periods. In the first period, suppose the best-fit equation is

$$q_1 = F_1(m_1, m_2, \ldots, m_N) \quad , \tag{3-4}$$

where $q_1$ is an air quality parameter, $m_1, \ldots m_N$ are meteorological parameters and $F_1$ is some function of those parameters. In the second period, suppose the best-fit equation is

$$q_2 = F_2(m_1, m_2, \ldots, m_N) \quad .$$

Then we examine the ratio of the equations:

$$\frac{q_2}{q_1} = \frac{F_2(m_1, \ldots, m_N)}{F_1(m_1, \ldots, m_N)} \quad . \tag{3-5}$$

This can be calculated for any set of meteorological conditions. If emission potential is unchanged, we would expect air quality to be about the

same in both periods, assuming that all significant meteorological parameters have been identified. Hence, the ratio $q_2/q_1$ will be about 1 for all meteorological conditions if the emission potential is unchanged. If $q_2/q_1$ is approximately constant over all meteorological conditions, but at a value other than 1, then emission potential has increased or declined by an amount given by that constant. (If the constant is 0.90, meteorologically adjusted air quality, i.e., emission potential, is 90% of its former value; that is, it has improved 10%.)

If $q_2/q_1$ is highly variable, then the improvement in emission potential depends on the meteorological condition, i.e., the meteorological parameters used do not adequately relate to the pollutant, and a more detailed examination is required. One might find, for example, that there is a degradation in adjusted air quality when the wind blows from the direction of a newly constructed power plant, but an improvement otherwise.

Mathematically, one could calculate $q_2/q_1$ by Eq. (3-5) for the meteorological conditions of each day during both periods. The median of all these values might be taken as representing the meteorologically adjusted change in air quality between the two periods. The 95th and 5th percentile values of $q_2/q_1$ might be used to indicate the range of uncertainty in this estimate.

Meteorological Adjustment for Multiple Time Periods

Sidik and Neustadter [33] proposed a linear regression to correct a series of annual pollution values for the Cleveland area. A linear regression of pollutant levels was performed against 29 predictor variables; the variables were meteorological, except for two rough indicators of economic

activity, a seasonality variable, and the number of days from the inception of the monitoring. Because this last variable increases linearly with time, it has the potential of accounting for a linear trend in emissions.

A similar method of meteorological adjustment was described by the EPA [41] using seasonally adjusted averages of wind speed, temperature, and rainfall for the Los Angeles area. Statistical analyses were performed with a nine-station composite of seasonally adjusted values of $SO_2$ and TSP. It was shown that $SO_2$ was correlated to inverse wind speed and that TSP was negatively correlated with rainfall. Observed quarterly levels were normalized to the average meteorological conditions for each calendar quarter. A variable corresponding to the year was used to model a linear emissions trend. Since only one meteorological variable was considered significant for each pollutant, that variable was used to adjust observed values using Eq. (3-2).

In the last two studies referenced, a "dummy" variable was used to represent a linear trend in emissions. This procedure deserves further discussion. In reference to Figure 3.1, we noted earlier that direct regression over several years could result in a coefficient which incorporated changes in emissions rather than meteorology alone. In Appendix A, we generalize this problem to the regression of daily values in several years. Our conclusion is that it is inadvisable to use data for several years in a regression against meteorological variables alone. Instead, one should either use dummy variables or do a separate regression in each year.

If dummy variables are used, we recommend using one for each year rather than a single linear trend variable. (See Appendix A for details.) If the

trend is not linear (e.g., is U-shaped), a linear trend variable can create misleading results.

If a separate regression for each year is performed, a method such as that recommended by Meisel and discussed above must be applied to each successive pair of years to obtain an adjusted trend.

Cautions Regarding Linear Regression

Because linear regression is one of the best understood and easiest to use statistical analysis techniques, it can be a powerful tool in meteorological adjustment, as discussed. Some cautions are in order, however; it is not always the best approach.

The most serious limitation of linear regression is that its effectiveness depends on the relationship between the meteorological variables and pollutant concentration being approximately linear. As we discussed in Section 2.0, many meteorological variables are nonlinearly related to pollutant concentrations. If nonlinearly related variables are avoided, only a portion of the variation in pollutant concentrations explained by meteorology will be modeled. This omission can potentially result in distortions: year-to-year differences in nonlinear variables would not be reflected in the adjusted values.

A nonlinear variable which is monotonically related to the pollution concentration will result in a correlation with that variable. The use of such a variable in a linear regression can result in biased residuals. (See Figure 3.2.) The resulting equation will predict high for some ranges of the meteorological variable and low for others. This condition can often

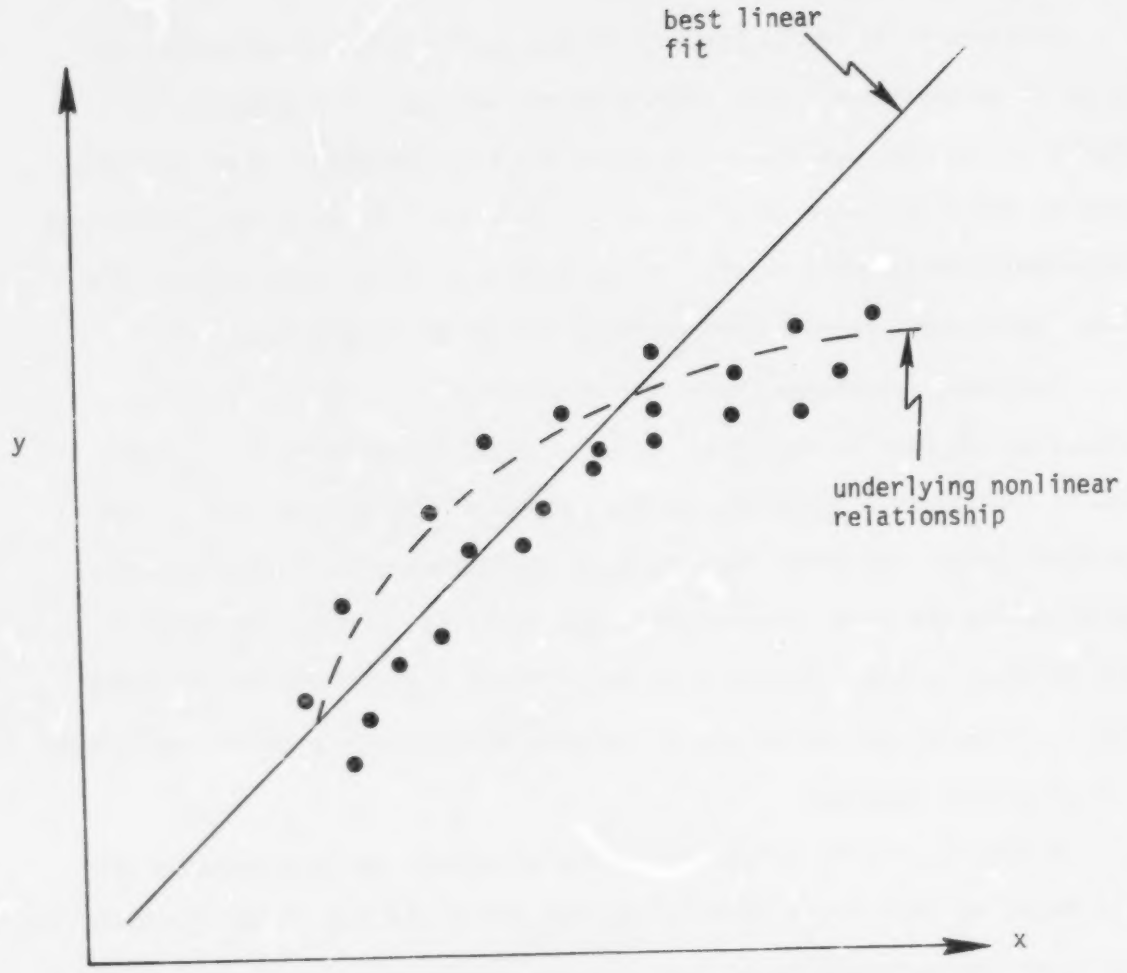


Figure 3.2 Example of Linear and Non-Linear Relationships. (The meteorological variable x is nonlinearly related to y, but fit with a linear equation. The linear equation will consistently overpredict for large values of x.)

be detected by plotting the residuals (difference between predicted and actual values of y) versus y. If the regression is legitimate, there should appear to be no relationship between the two.

Nonlinearities can be handled through redefinition of variables, in order to maintain the linear regression methodology. For example, it is common to use the logarithms of a meteorological variable. More generally, tabular definitions are possible, as in Table 3.1. In this case, Zeldin has transformed nonlinearly related variables into a linear point score. One then simply uses the redefined variable in a linear regression.

Nonlinear transformation of the concentration variable y are also possible, but care is required. If we predict, for example, log y with 40% of the variance explained, we are not explaining the same 40% of the variance in y. The predictions must be transformed into predictions of y to calculate the error in predicting that quantity. Explaining more of the variance in log y than in y is not sufficient justification for using log y. Instead, the percentage of variance explained in y by the log y regression should be computed.

On the other hand, since most pollution values can be reasonably approximated as lognormally distributed, the use of the log of the concentration may cause the residuals to be less biased. The equations in this section and in Appendix A are valid if y is interpreted as the log of the pollutant concentration.

Another typical error in linear regression is to use too many variables. One can overfit the data by adding variables which do not significantly improve that fit; this practice can cause anomalous results. A

further difficulty in using too many variables is that dependencies among the variables can make the coefficients uninterpretable. If linear regression is used, we recommend stepwise linear regression to avoid using anymore variables than necessary.

### Nonlinear Regression

In general, if a relationship is nonlinear, it is better to perform nonlinear regression directly rather than use a number of tricks to make the problem linear. However, software for nonlinear regression is not as common or automatic as linear regression. Techniques of the following sub-section, however, allow nonlinear effects to be incorporated without nonlinear regression.

## 3.3 CLASSIFICATION APPROACHES TO METEOROLOGICAL ADJUSTMENT

The basic difficulty with the regression approach is finding a simple relationship between meteorology and air quality which holds for the full range of meteorological variables. The approach of this subsection is to reduce this difficulty by breaking up the problem. In particular, we focus attention on days falling in meteorological classes which are conducive to high pollution levels.

Such a method of adjusting oxidant data in the San Francisco area was employed by the Bay Area Air Pollution Control District [1]. By taking two meteorological parameters known to have a significant relationship to oxidant levels on a daily basis, adverse days were selected for seven locations. For each location, a limiting temperature and an inversion base height were used to classify days into the "adverse" category. Annual averages of the daily

maximum hourly concentrations were computed based on only those days classified as "adverse." The averages from year-to-year thus represented the meteorologically adjusted oxidant trend.

The applicability of this technique is apparent. A specific subset of days containing similar meteorological conditions approaches the ideal situation of holding meteorology constant. The trend resulting from the analyses can be regarded as a good representation of the downward oxidant trend due to effectiveness of emission control strategies.

One potential problem, however, in employing this method is the variability of meteorological conditions within the prescribed subset of days. If the criteria selected are too broad, the variability within categorized subsets may not adequately describe the desired "adverse" effect. In essence, meteorology is not held constant, and ultimately, results cannot be regarded with confidence. On the other hand, criteria set too stringently can reduce the number of such days to the point of losing statistical confidence in the results. Appropriate class definition is critical. Again, even with good definitions, very large deviations from normal weather patterns can seriously impact the accuracy of this model. (See Appendix A.)

A method for weather-correcting oxidant data in the San Gabriel Valley portion of the Los Angeles Basin was developed by Kerr [24]. He classified September meteorological data for the years 1968 to 1973 into five meteorological types: (1) closed high-pressure center aloft, (2) inversion < 1100 ft., (3) inversion 1100-3500 ft., (4) surface Santa Ana wind type, and (5) closed low or deep trough aloft. These definitions involved some subjectivity and

were not mathematically explicit. A table was constructed depicting a frequency distribution by weather category per year. For each category, multi-year averages of the maximum hourly averages were determined.

Once the distribution of days by category was established, the weighted averages of each year of record were computed; that is, Kerr summed the yearly frequency of occurrence over all the categories and multiplied by the concentration mean for each category. Thus, the expected concentration value for a year, given the meteorology of that year, could be calculated. The residuals reflecting the difference between this predicted value and the observed value were used as a measure of meteorologically adjusted trends.

## 3.4 METEOROLOGICALLY INVARIANT STATISTICS

Another method, presented by Zeldin et al.[46], explores the possibility of performing meteorological adjustment to air quality data without the use of meteorological data!

In this particular study, oxidant data were analyzed in the Los Angeles area. It is known that meteorological variability affects annual ozone concentrations. However, even in the most favorable of meteorological years, there will still occur a number of particularly adverse days under which elevated ozone concentrations will occur. The author found that, by using the fifth-highest maximum hourly average at a given site for a particular year, and then comparing that value to the same parameter in other years, the general meteorological patterns under which the ozone concentration occurred were nearly identical. Thus no specific meteorological parameters were used, but the values reflect underlying emission trends.

This method is basically untested. It is not known whether the methodology can be utilized more universally; however, it does offer potential use in areas where meteorological data are sparse. If proven to be an effective method of providing a meteorological adjustment of air quality data, the method allows one major advantage -- simplicity. Since this approach, strictly speaking, exploits meteorological invariance rather than meteorological adjustment, we will not pursue it further here.

# 4.0 RECOMMENDED PROCEDURES FOR METEOROLOGICAL ADJUSTMENT OF AIR QUALITY

This section describes in detail a recommended methodology to achieve a meteorological adjustment of air quality data, which in turn can result in an air quality trend based on meteorologically adjusted emissions, i.e., emission potential. Guidelines are presented for the definition of distinct meteorological classes and for the use of these classes to determine a meteorologically adjusted trend. By determining average concentrations for each class, and then by relating the data to a "typical" meteorological year, a pollutant concentration trend can be obtained which provides an estimate of what would have occurred if all years had closely comparable meteorology.

It should be noted that these recommended procedures, while based on extensions of ideas generated previously in the literature, have not been tested. Thus, until adequate evaluation programs have been undertaken, results of these recommended procedures should be qualified.

## 4.1 SELECTION OF METEOROLOGICAL CLASSES

In Section 2.0, we discussed meteorological parameters which affect given pollutants and the nature of those effects. Presuming that one has a list of candidate meteorological variables for defining meteorological classes, it remains to define the classes using all or some of these candidate variables.

A "meteorological class" is a set of days which have a specific set of meteorological characteristics. Such classes can be naturally occurring

weather categories such as those listed in Table 4.1. In this type of classification process it is quite possible for two weather classes to have the same average pollution level, i.e., the same "pollution potential." As in Table 4.1, classes 1 and 5, although meteorologically dissimilar, both result in low pollutant concentrations. It is clearly unnecessary, therefore, to be concerned with fine distinctions between meteorological classes which always result in low pollution concentrations.

The simplest categorization possible is into two classes: the "adverse" class and the "non-adverse" class, as in the Bay Area APCD work [1] reviewed in Section 3.3. The adverse-class approach has the disadvantage of requiring a very accurate definition of meteorologically adverse days. If the definition misses many high-pollution days or includes too many low-pollution days, the resulting trend estimate may be unreliable. If a greater number of classes are defined, the requirement for precision in defining each class is reduced.

The definition of meteorological classes may depend to some extent on the use made of the results of the analysis. It is possible to have distinctly different trends in each meteorological class if the major emission sources differ for each class (due to wind direction, for example). Even if two classes which are meteorologically distinct have approximately the same average pollution levels, one may wish to preserve the distinction; doing so may allow isolating the effect on the overall trend of specific emission sources.

Table 4.1 Meteorological Classes Established by Weather Type for the Southern California Area [27].

| Class | General Weather | Associated Synoptic Pattern | |
|---|---|---|---|
| | | Surface | 500 mb |
| 1 | Hot, dry continental air throughout the day ("Santa Ana" conditions) | Large high pressure over Great Basin | Strong northerly winds over area with lower pressure east of the area |
| 2 | Relatively dry forenoon, modified marine air in afternoon, very hot (heat wave) | High pressure over Great Basin and thermal trough over desert | Sub-tropical high over area |
| 3 | Moist, modified marine air; hot in afternoon | Thermal trough over desert | High pressure over area |
| 4 | Moist, modified marine air; warm in afternoon | Thermal trough over desert | Trough over area |
| 5 | Cool, moist, deep marine air throughout the day | Synoptic-type low over desert | Deep trough or closed low over area |

There are two basic ways to define meteorological classes, differing largely in emphasis:

(1) Define the classes through experience and meteorological expertise, guided by examination of historical data; or

(2) Define the classes using a data-directed empirical procedure, guided by experience and meteorological expertise.

We examine these two approaches in the following two subsections.

4.1.1 Definition Through Experience and Expertise

The experience gained in working with meteorological and air quality data can serve as a basis for defining categories.

If a meteorological pollution-potential index is defined from basic principles, as exemplified in Section 3.1, classes can be defined by use of the index. The range of values taken by the index can be divided into intervals; each interval then corresponds to a meteorological class. For example, using Zeldin's point scoring system (Table 3.1), a given day has a score between 0 and 40, based on the values of four meteorological variables. Days with a score over 34 might form one class, days with a score between 29 and 34 another class, and so on. When classes are defined this way, they should all have different average pollution levels.

Classes might be defined using knowledge of obvious meteorological classes common to a given area, as was illustrated in Table 4.1. Another example is "Rule 57" days, discussed in Section 3.1. If, as in Table 4.1, the class definitions are not mathematically explicit, the analyst must be prepared to classify every day used in the analysis by inspection of the meteorological data.

If the subjective approach to defining classes is taken, it is important to test if there are significant differences in average pollution levels between at least some classes; otherwise, the categorization is inappropriate for trend adjustment. Thus, air quality data are used to check the classes defined.

4.1.2 Empirical Derivation of Meteorological Classes

Because the cause-effect relationships between meteorology and air pollution are sometimes counter-intuitive, it may be necessary to use data to guide the definition of meteorological classes more explicitly. The data required are the daily meteorological variables and the corresponding daily pollutant values.*

Some computationally demanding methods of deriving such classes could be employed. For example, a program called AID, developed at the University of Michigan, can be used to analyze meteorological data and corresponding pollution levels [34]. It will generate from such data definitions of meteorological classes so that the pollutant levels on days within each class are as uniform as possible and such that the levels in each class are as distinct as possible. Other approaches to this class-definition problem are available in the literature [49], but not as standard statistical packages.

In Appendix B, we describe a procedure which uses common statistical software packages (for example, SPSS and BMD) and human judgment to implement a powerful approach to the data-directed derivation of meteorological classes which discriminate pollution levels. To illustrate the basic approach, we will use as an example work by Davidson [14].

---

*The daily value chosen is usually related to a standard, e.g., the daily maximum one-hour-average oxidant concentration.

Davidson attempted to define meteorological classes in Los Angeles which corresponded to different oxidant concentrations. He used data to guide his definition of those classes. He plotted scatterplots of one meteorological variable versus another. On such a plot, the values of the two variables for a given day define a point. He identified the pollution potential on that day by indicating that point on the plot which corresponded to an alert day (maximum one-hour average oxidant above 35 pphm) or a non-alert day. In Figure 4.1, an example of an alert day ("A") and a non-alert day ("N") is indicated. By examination of such a plot, regions corresponding largely to alert days can be "fenced off"; similarly, regions corresponding to non-alert days can be defined. In Figure 4.1, regions C and A are directly defined; that is, given a day with particular values of the meteorological variables, the region to which it corresponds can be determined from the plot.

Area B in Figure 4.1 does not have a clear character, having many alert and non-alert days. Thus, when a day fell in region B, Davidson used other meteorological variables to further differentiate meteorological classes such that every day could be explicitly classified. His full classification scheme is shown in Figure 4.2.

This procedure can be represented as a decision tree (Figure 4.3). Such an end result has a number of advantages. The most important is that it allows one or two variables to be used at each "decision box" in the tree, so that visual derivation of the boundaries is possible; yet a larger number of variables can be used in the overall classification. It also allows the natural incorporation of subjective judgment and experience in guiding the choice of boundaries and variables.

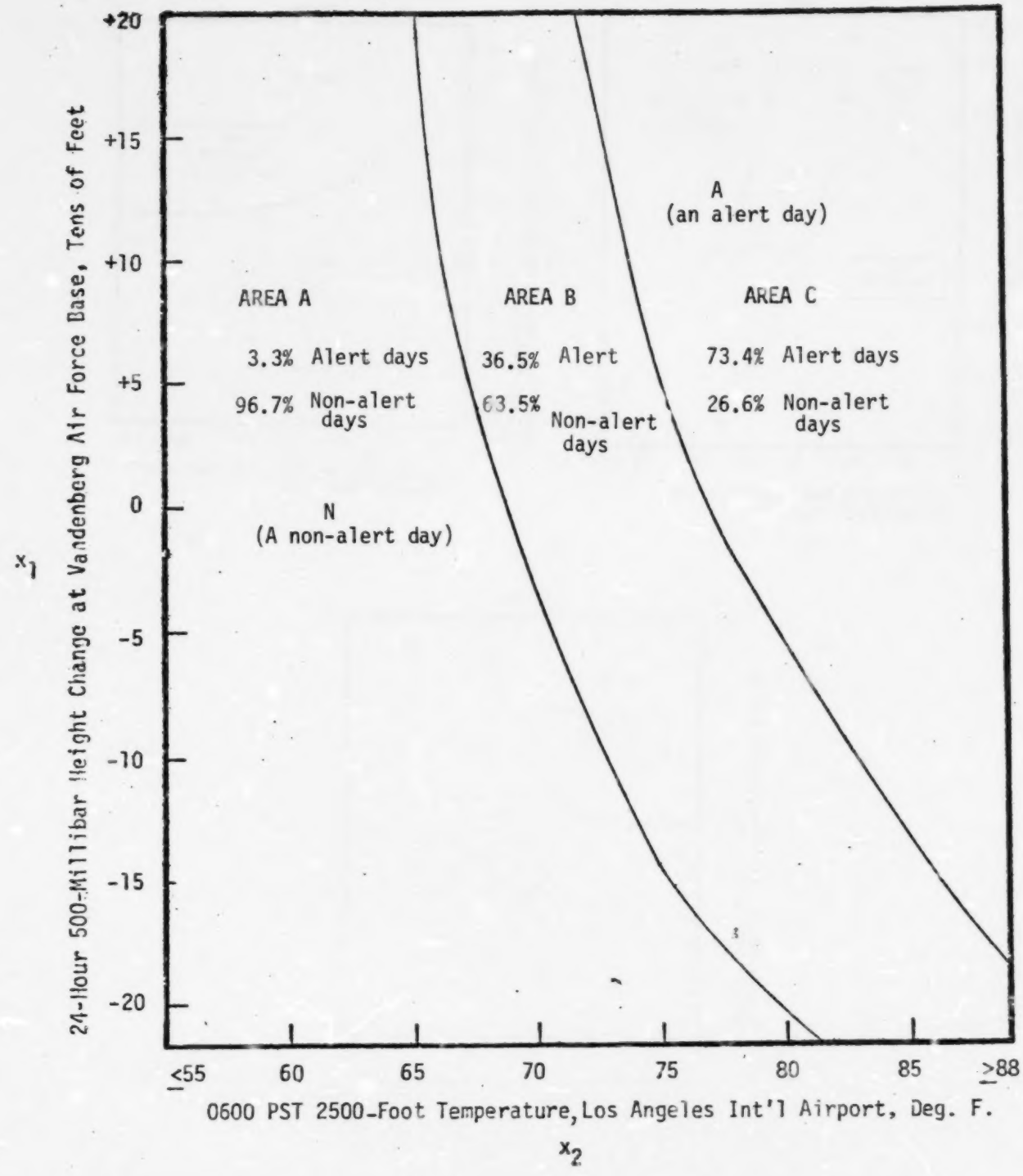


Figure 4.1 First Stage of Discrimination Employed by Davidson.

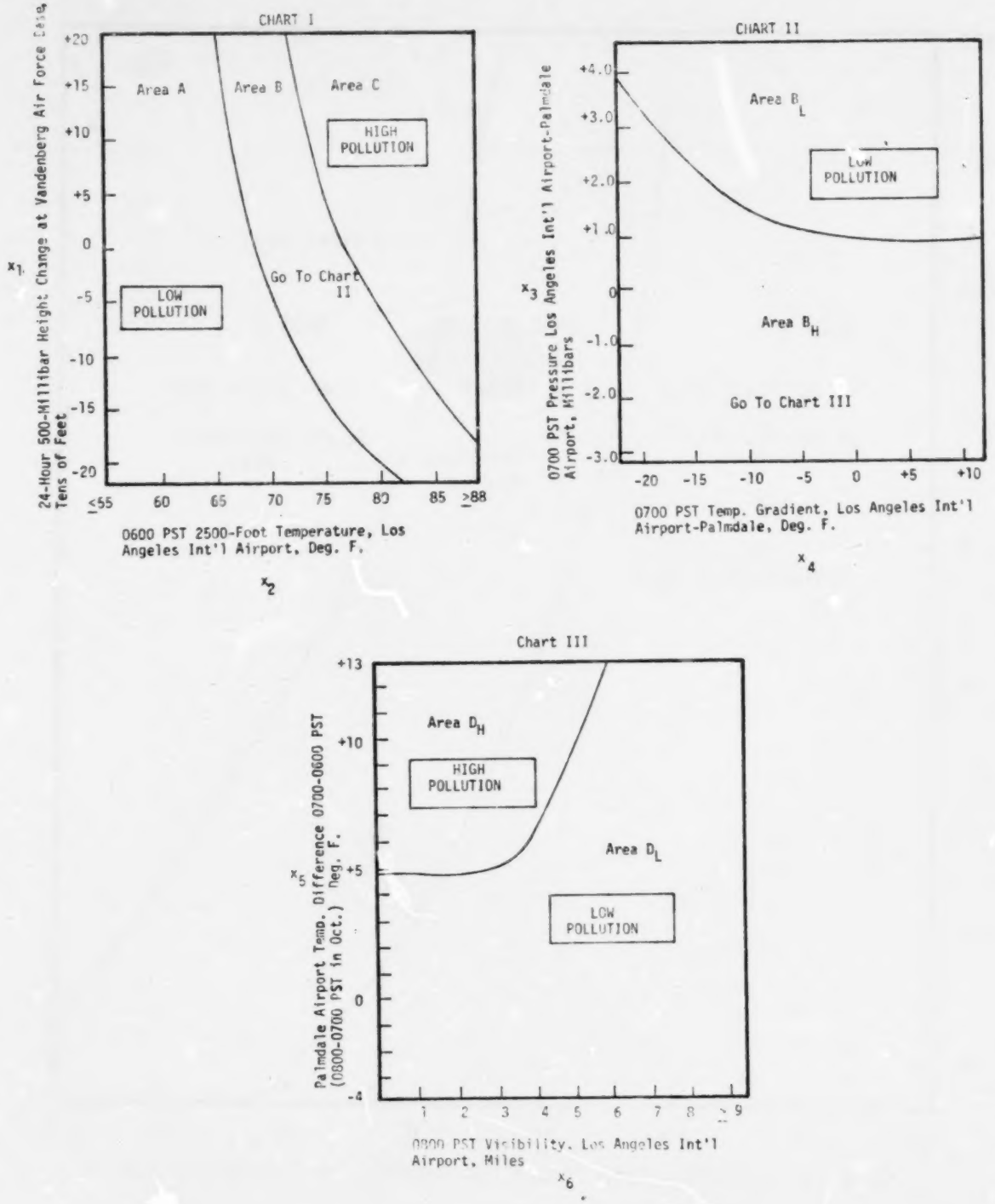


Figure 4.2 Full Prediction Scheme by Davidson. The prediction for each area is indicated.

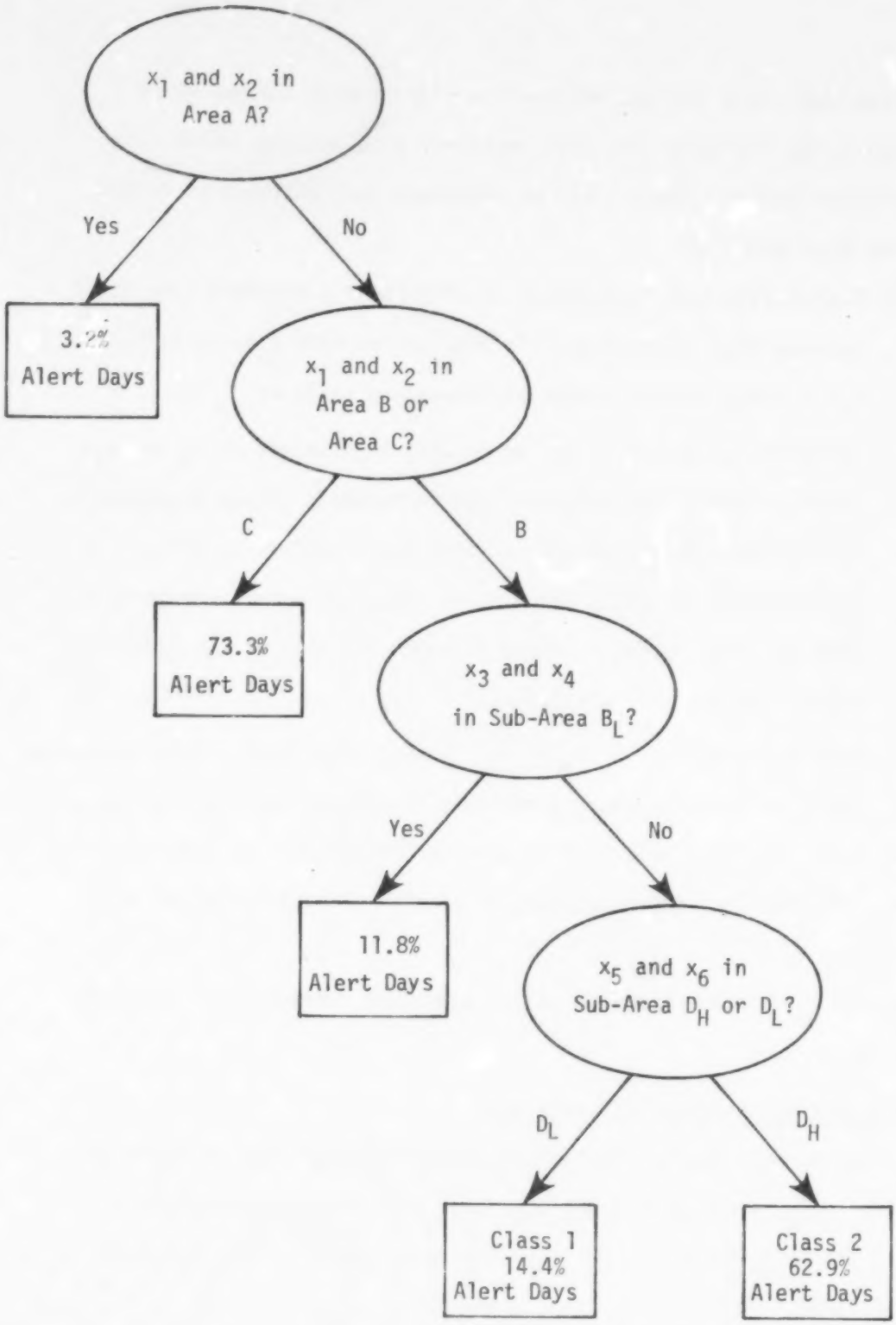


Figure 4.3 The Decision Scheme of Figure 4.2 Drawn as a Decision Tree.

Notice that there are two meteorologically adverse classes which correspond to two different "natural" meteorological classes which cause high pollution levels. These could be described qualitatively as follows (referring to Figure 4.2):

(1) A high 2500-foot temperature at 0600 Pacific Standard Time at LA International Airport ($x_1$) in combination with a large 24-hour 500-millibar height change at Vandenberg ($x_2$); or

(2) Intermediate values of $x_2$, or jointly high values of $x_1$ and $x_2$; low-to-intermediate values of the LA-Palmdale pressure gradient at 0700 PST (or low values of temperature difference between LA and Palmdale at 0700); and a large temperature increase between 0600 and 0700 PST at Palmdale in combination with low visibility at LA International Airport at 0800.

One might treat the combination of the two classes as a single class ("adverse days") or treat the two subclasses separately in determining trend by methods discussed later. In fact, all five classes resulting from the tree could be used. We recommend that the subclasses be treated separately rather than grouped.

Appendix B elaborates in some detail a similar procedure for deriving such a tree.

## 4.2 A METEOROLOGICAL ADJUSTMENT PROCEDURE

Once the appropriate meteorological parameters have been selected and the specific meteorological classes have been determined, a procedure to adjust a multi-year set of air quality data for meteorological variation must be utilized.

A meteorological adjustment can be performed rather simply by using a "typical-year" approach. This can be accomplished by determining the average frequency of occurrence for each usable category over the analysis period (e.g., 5 years). These values represent the "average" occurrences, thereby delineating the basis for a "typical" meteorological year. From the typical-year frequencies, observed concentrations by category and year can be used to calculate the concentration values for each year which would have occurred if the year had typical meteorology. In other words, average concentrations are restated for each year assuming each year has the same meteorology. This results in normalized values which relate to the same scale of values represented by the actual observations. Thus actual trend data and the meteorologically adjusted trend data can be plotted in a superimposed manner to achieve an easy-to-interpret comparison, the adjusted values showing the underlying trend in emission potential and the difference indicating the effect of favorable or adverse meteorology. It also becomes meaningful to calculate meteorologically adjusted percentage improvements.

To achieve this objective, we proceed as follows:

Tabulate

$n_{km}$ = the number of days in meteorological class k in year m

$\overline{C}_{km}$ = the average concentration over the $n_{km}$ days in class k in year m

M = the number of years under consideration

K = the number of meteorological classes used

$$N_k = \sum_{m=1}^{M} n_{km}$$

(the number of days in class k in the entire M-year period)

$$N = \sum_{k=1}^{K} N_k$$

(the total number of days in all K classes and all M years).

Then

$$\overline{C}_m = \sum_{k=1}^{K} n_{km}\, \overline{C}_{km} \Big/ \sum_{k=1}^{K} n_{km} \qquad (4\text{-}1)$$

is the observed average concentration in all K classes for year m. These values for each year represent a measure of the unadjusted trend.

The quantity

$$P_k = N_k/N \qquad (4\text{-}2)$$

is the "typical" fraction of class k days occurring in a year (based upon the frequency of occurrence of class k days over the M-year period). Then a "typical" year will be one where days in the $k^{th}$ class occur with a frequency $P_k$.

The average pollutant concentration over all K classes in year m if the distribution of meteorological classes was "typical" is then

$$\hat{C}_m = \sum_{k=1}^{K} P_k\, \overline{C}_{km} \quad . \qquad (4\text{-}3)$$

These values represent the meteorologically adjusted trend, i.e., the average concentration if each year had the same distribution of meteorological classes.

Determination of "Typical-Year" Conditions (for Days Exceeding Threshold)

A method similar to that of concentration normalization can be employed to determine "typical-year" conditions for the number of days exceeding some

designated threshold level (e.g., the NAAQS). Using the same category structure as constructed for the previous analysis, computations can be based on the percentage of days exceeding the threshold value by category.

If we let $E_{km}$ be the number of exceedences of the threshold level for category k in year m, then $\overline{C}_{km}$ can be replaced by

$$\overline{E}_{km} = \frac{E_{km}}{n_{km}} ,$$

the fractional number of exceedences for category k in year m. Following the procedures outlined previously, we compute

$$\hat{E}_m = \sum_{k=1}^{K} \overline{E}_{km} P_k ,$$

which now represents the meteorologically adjusted fraction of days exceeding the threshold level in year m. To obtain the meteorologically adjusted number of days per year exceeding the threshold, we just multiply $\hat{E}_m$ by 365.

Limitations on Averaging Period

We have discussed the methodology as an average of daily values over the entire year. Naturally there will be, in any practical situation, days for which all data are missing or for which sufficient data are missing to make the estimate of the standard-related value (such as the maximum 1-hour average) suspicious. Further, some pollutants such as TSP are typically sampled rather infrequently (e.g., every 6th day). There is nothing in the methodology developed which changes its validity if the analysis is limited to days of the year in which data are available and the averages are taken over those days of the year. When the missing data are randomly spaced throughout the year,

the probabilities of each meteorological class occurring should be determined from as full a set of meteorological data as possible and not be limited to those days on which pollution data are available; however, the average concentration within each class can (and indeed must) be limited to those days in that class on which data are available.

The same logic applies to pollutants for which there is a strong seasonal effect. For example, high oxidant concentrations are centered about the late summer, and carbon monoxide concentrations are higher in the months of October through February. One could thus look at the trend from year to year in the pollution season. In that case, the averaging is again done over the available data in the season to which the analysis is restricted. In this case, however, the relative frequency of meteorological classes should be determined only for meteorological data in the seasons being analyzed.

Limitations of Data Availability

A minimum of five years of data is recommended to determine a reasonable "typical year." Once this has been established, continuing trend analyses can be performed in reference to the already established typical-year data. If, on the other hand, it appears, at some future time, that the typical-year meteorology of the original data set was no longer representative of the longer-term period, then the $P_k$'s can be re-computed to establish a new typical-year data base. The recommendation for five years of meteorological data does not necessarily imply that five years of pollutant data are required.

Number of Categories Used

In some cases we may be interested only in trends of the highest pollution levels. A possible procedure is to limit analysis to those categories comprising approximately 50% of the sample days containing the higher concentration. For example, if we classify the days into six categories, and if half the days occurred in high-pollution classes 1 through 3 while the other half occurred in low-pollution classes 4 through 6, then the latter group would be excluded from further consideration. If this approach is taken, the formulas of this subsection can be interpreted to be over the K worst meteorological classes.

Caution, however, must be exercised if the classes used are not inclusive of all days in the year. If 50% of the days are excluded at one locale and 60% at another, the two trends may not be strictly comparable. If trends are stronger for the more adverse classes, then including more lower-pollution days will make either an uptrend or downtrend less pronounced. Thus, if trends at several locales are to be compared, it is perhaps best to use classes which encompass all days, and perhaps to note trends in individual meteorological classes as well.

A way to minimize (and possibly eliminate) the uncertainty caused by differing absolute trends in each class is to average the logarithms of the concentration; i.e., wherever we have used $\overline{C}_{km}$ in the equations, $\overline{\log C_{km}}$ can be substituted. Then, the year-to-year change in adjusted pollutant levels is related to the percentage changes in each category rather than of the absolute changes in each category.

Adjustment of Means within Classes

There is an implicit assumption in the methodology discussed that each meteorological class is sufficiently well-defined that the meteorological

conditions occurring from year to year do not vary significantly within the class. If they do, then one can meteorologically adjust the average concentration within each class. Since this is a second-order effect, it does not deserve extensive discussion. We, however, note that the linear regression approach can be used to adjust the mean within a class. (It is much more reasonable to assume that a linear relationship holds within a class than for all days.)

Given a derived regression equation for data within a class (such as shown in Appendix A, Equation (A-3)), how should it best be used for meteorological adjustment of observed values? We can extend Equation (3-2) (which related to a single variable) to multiple variables. Let

$\bar{x}_i$ = the average value of the $i^{th}$ meteorological variable over all years of interest (within the class)

$\bar{x}_i^k$ = the average value of the $i^{th}$ variable in the $k^{th}$ year (within the class)

$\bar{y}^k$ = the observed average of daily pollution values over year k (within the class).

Then the meteorologically adjusted values for each year (for the class being adjusted) are

$$\hat{y}_k = \bar{y}^k + \sum_{i=1}^{N} a_i \, (\bar{x}_i - \bar{x}_i^k) \quad .$$

The technique discussed in Appendix A should be used to derive the regression coefficients for each meteorological class separately.

## 4.3 CHOICE OF THE AIR QUALITY MEASURE TO BE ADJUSTED

We have noted previously that one element of meteorological adjustment procedure is the basic decision of which air quality measure (or index) to adjust. We have delayed a specific discussion of the tradeoffs involved in this choice to this point in the report in order to place it in the perspective of available methodology. (Some general considerations were addressed in Section 2.2.2 on the differences in meteorological effects on various air quality parameters.) In our recommended procedure, we settled on two measures:

(1) The annual average of daily standard-related values (daily 1-hour maximum, daily 8-hour maximum, daily average); and

(2) The fraction of days in the year exceeding a pre-specified threshold (e.g., the fraction of days each year the daily maximum 1-hour average oxidant exceeded the secondary standard).

We will discuss the motivation for these particular choices of air quality measures and then discuss other alternatives.

Since the daily standard-related value (e.g., the daily 1-hour maximum) is a measure of the health effect on that given day, then the average of daily values yields an overall average impact for the year or for the portion of the year where the problem exists. This is a useful measure of overall progress in pollution control for all meteorological classes and for all pollutant levels; it is, however, not directly related to progress in meeting the air quality standards. These standards emphasize the more meteorologically adverse (high pollution-potential) days.

The second measure, the number of days per year for which the pollutant concentrations exceed a specified value, is a measure oriented directly toward progress in meeting the standards. It can be derived separately for primary and secondary standards and can also be limited to the pollutant season. A meteorologically adjusted trend in the number of days in which the standard is exceeded is a very direct measure of progress in meeting the standards, yet it is not quite so particular to a single day's meteorology as is the annual maximum of the standard-related value.

An adaptation of the first measure which makes it similar to the second measure was suggested in Section 3.0--limitation of the meteorological classes considered to the more adverse classes. This limitation is more or less automatic with the second measure: Meteorological classes with no exceedences will not be used and thus need not be defined.

Other Alternatives

Many standards are stated in terms of the annual maximum of a short-term average, and these standard-related values are regularly reported. One could ask for the meteorologically adjusted value of such a quantity. The difficulty in doing so directly is that such values often occur due to unusual and extreme meteorology and are not easily amenable to adjustment. It would, for example, be difficult to identify a "typical" worst case to which one might adjust.

However, the measures of the annual average standard-related value and the exceedences of a threshold value can be combined to give an estimate of the meteorologically adjusted standard-related value. Specifically, suppose

we are interested in the annual maximum 1-hour average of a pollutant. Suppose we calculate the meteorologically adjusted annual mean of the daily maximum 1-hour average, say $\mu$. Suppose we have calculated the meteorologically adjusted probability, P, of the 1-hour maximum in a day exceeding the threshold, $\tau$. Then we have two parameters of the meteorologically adjusted distribution of the daily maximum 1-hour average: The mean $\mu$ and the $(1-P)^{th}$ percentile value $\tau$. If we assume a two-parameter distribution for the pollutant, e.g., log-normal, then the distribution is fully determined. Any percentile of the distribution, including the percentile corresponding to the annual maximum, can be estimated. (The highest day of 365 can be estimated as the $99.7^{th}$ percentile.)

Another alternative in measuring the pollution trend is the use of other than daily measures. One might propose to adjust, for example, monthly, quarterly, or annual data. We have discussed approaches to doing so which adjust such longer-time-period measures by the number of adverse days in the period or by the value of a meteorological parameter or index. When daily values are used, the meteorology for the day is directly related to the pollution level for the day. When data are aggregated over a larger time period, the direct correspondence is lost.

One could further <u>disaggregate</u> and go to hourly values rather than daily values. If the methodology were extended in this direction, the implication would be that hourly meteorology is related to hourly pollutant values, an assumption seldom satisfied. Further, hourly values of pollutant and meteorological measurements are certainly not independent, while day-to-day levels of pollution and meteorological variables are considerably more independent.

## 4.4 AN EXAMPLE OF METEOROLOGICAL ADJUSTMENT OF AIR QUALITY DATA

This subsection presents a simple hypothetical example to illustrate the type of trend analysis possible and to provide numerical examples.

### $SO_2$ Point Source to Receptor

In this hypothetical situation, a point source of $SO_2$ emissions is located 10 miles due east of an air monitoring facility. Over a five-year period, control efforts have supposedly reduced $SO_2$ emissions, but increased workload demands upon the facility have resulted in an increase in production rates. The problem now is to determine which of the two counter-balancing measures has had the greatest impact upon the ambient $SO_2$ air quality. Examination of the annual average $SO_2$ concentration levels from the air monitoring facility has been inconclusive due to the year-to-year meteorological variations (see Table 4.2).

Table 4.2 Annual $SO_2$ Average Concentrations for the Hypothetical Receptor Site

| | YEAR | | | | |
|---|---|---|---|---|---|
| | 1968 | 1969 | 1970 | 1971 | 1972 |
| Annual $SO_2$ Average Concentration (PPM) | .082 | .088 | .100 | .077 | .089 |

A search for meteorological data indicates that wind records are available at both the source (S) and receptor (R) locations. Using 1968-1972 data, mutually exclusive categories are selected as follows, based upon the severity of wind conditions causing transport of the $SO_2$ emissions from S to R:

Category 1: Wind direction at both S and R are from the east for at least 12 hours per day.

Category 2: Wind direction at both S and R is from the east for at least 4 to 11 hours per day.

Category 3: Wind direction at either S or R, but not both, contains an easterly component for at least 4 hours per day.

Category 4: Wind direction at either S or R, or both, contains an easterly component for 1 to 4 hours per day.

Category 5: Wind directions at both S and R have no easterly component during the entire day.

For the five years of analysis, the average daily $SO_2$ concentration is evaluated against the wind data as specified in the five categories. Results are shown in Table 4.3.

Assume we are interested only in the most adverse 50% of the days. The category cut-off point is determined by summing the $P_k^{(1)}$ values until the top categories approximate 0.50. In this instance, in categories 1 through 3, $\Sigma P_k^{(1)} = 0.487$. Therefore, categories 4 and 5 will not be processed any further.

The next step in the process is to compute the mean concentrations for "typical-year" meteorology. Table 4.4 gives the tabular layout for the computation of the "typical-year" values.

Table 4.3 Tabular Layout of the $n_{km}/\bar{C}_{km}$ Entries and the A Priori Probability, $P_k$ for the Hypothetical Case

| CATEGORY | Year m | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n_{km}/\bar{C}_{km}$ | 1968 | 1969 | 1970 | 1971 | 1972 | $N_k/\bar{C}_k$ | $P_k^*(1)$ | $P_k^{**}(2)$ |
| 1 | 30/.203 | 14/.216 | 12/.254 | 5/.233 | 18/.199 | 79/.214 | .046 | .094 |
| 2 | 37/.156 | 21/.148 | 63/.183 | 20/.176 | 41/.166 | 188/.069 | .106 | .217 |
| 3 | 103/.077 | 135/.105 | 118/.096 | 108/.100 | 115/.098 | 579/.096 | .335 | .689 |
| 4 | 126/.059 | 97/.069 | 105/.060 | 161/.053 | 111/.057 | 600/.059 | .349 | |
| 5 | 61/.033 | 76/.041 | 41/.044 | 47/.045 | 55/.040 | 280/.040 | .162 | |
| $\frac{N_m}{\bar{C}_m}$ | 357/.082 | 343/.088 | 339/.100 | 341/.077 | 340/.089 | 1726/.087 | | |

$N_k = \sum_m n_{km}$ , $N_m = \sum_k n_{km}$

$^*P_k(1)$ is the a priori probability based on all categories $(= N_k/\Sigma N_k)$ .

$^{**}P_k(2)$ is the a priori probability using only the first three categories.

Table 4.4 Tabular Entries Showing the Product of the Top-Three-Category A Priori Probabilities, $P_k^{(2)}$, and the Average Category Concentration, $\overline{C_{km}}$. ($\hat{C}_m$ is the sum of the individual category entries and represents the normalized concentration by year.)

| CATEGORY \ YEAR | 1968 | 1969 | 1970 | 1971 | 1972 |
|---|---|---|---|---|---|
| 1 | .019 | .020 | .024 | .022 | .018 |
| 2 | .033 | .032 | .038 | .038 | .036 |
| 3 | .053 | .072 | .064 | .069 | .067 |
| $\hat{C}_m$ | .105 | .124 | .126 | .129 | .121 |

For comparative purposes, the same procedure for meteorological normalization has been computed for the full five categories. Results are shown in Table 4.5.

Table 4.5 Same as Table 4.4, except over all five categories, using $P_k(1)$, instead of $P_k^{(2)}$.

| | 1968 | 1969 | 1970 | 1971 | 1972 |
|---|---|---|---|---|---|
| 1 | .009 | .010 | .012 | .011 | .010 |
| 2 | .017 | .016 | .019 | .019 | .018 |
| 3 | .026 | .035 | .032 | .034 | .033 |
| 4 | .021 | .024 | .021 | .018 | .020 |
| 5 | .005 | .007 | .007 | .007 | .006 |
| $\hat{C}_m$ | .078 | .092 | .091 | .089 | .087 |

Figure 4.4 shows the actual and meteorologically adjusted trend data. The adjusted trend over categories 1 to 3 indicated that $SO_2$ conditions deteriorated until 1971, after which improvement began. One can see that the effects of increased production appear to have overshadowed the control efforts in the first few years, but that the control efforts did indeed have an effect upon improving the nearby air quality. Notice, too, that this conclusion is not discernible from the actual data.

Also shown in Figure 4.4 is the annual adjusted and unadjusted trend over all categories. In this analysis, the worst condition occurred in 1969, with slow improvement thereafter. The importance of this effect can be attributed to the impact of remote sources affecting background levels. Since the inclusion of categories 4 and 5 contain meteorological conditions not conducive to transport from the primary source facility, the change in trend indication is due to the impact from other sources being transported over much larger distances. Therefore, one can conclude that an overall improvement in background levels since 1969, due most likely to effective control efforts at the distant sources, has made a significant impact on the air quality at the monitoring location. Even though the shape of the unadjusted curves for both analyses are about the same, the meteorological adjustment methodology is able to distinguish both the effects from the primary source of $SO_2$ and the remote transport affecting background levels.

Let us now consider the evaluation of the trend of $SO_2$ NAAQS violations (0.14 ppm for a 24-hour average). An examination of the data from the

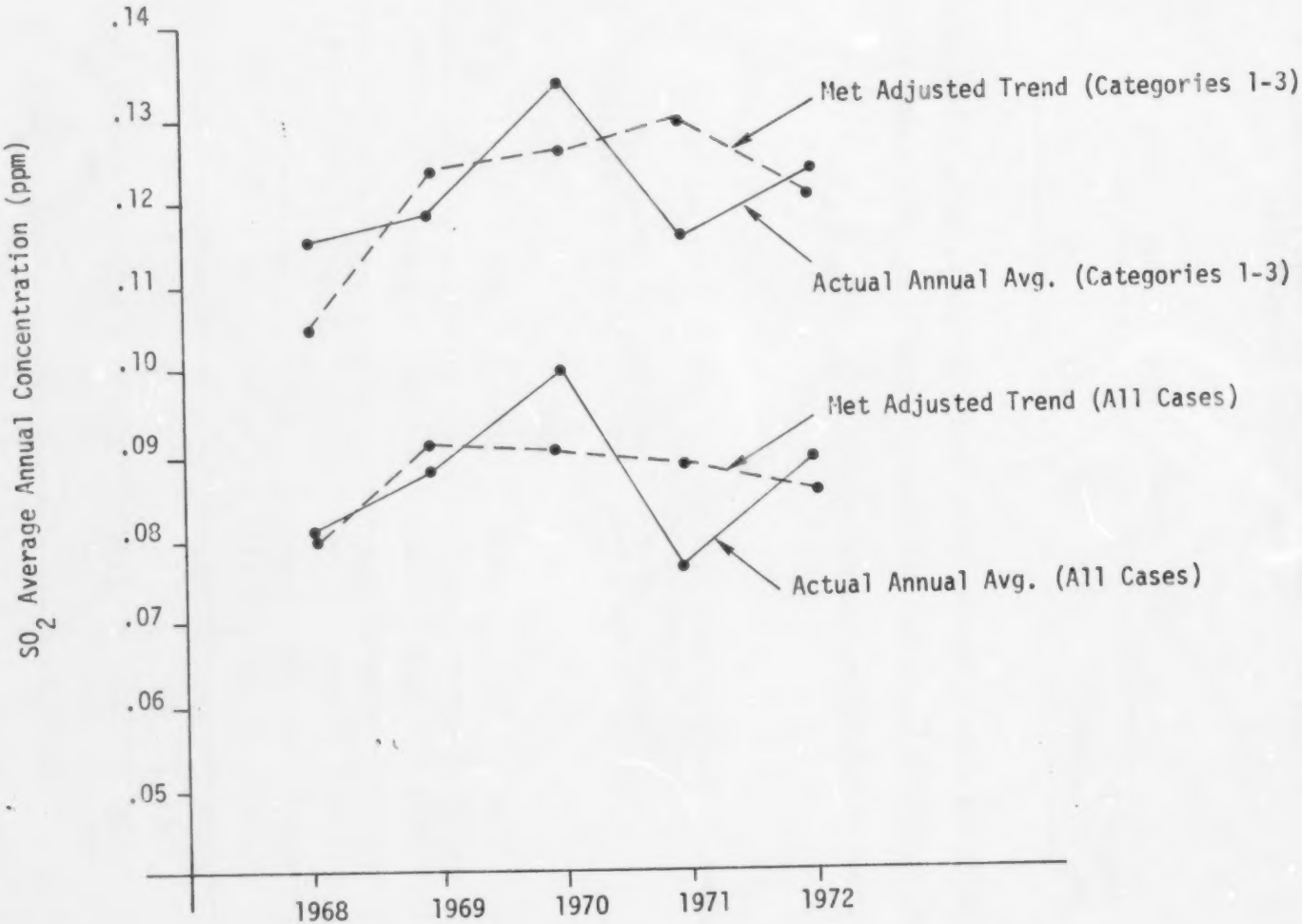


Figure 4.4 Results of Hypothetical Example Showing Actual $SO_2$ Trends With Associated Meteorologically Adjusted Trends.

receptor location indicates the following number of violations per year:

Table 4.6 Number of Annual $SO_2$ NAAQS Violations for Hypothetical Receptor Site

| | 1968 | 1969 | 1970 | 1971 | 1972 |
|---|---|---|---|---|---|
| Number of $SO_2$ violations | 62 | 48 | 76 | 32 | 57 |

Using the five categories established previously, Table 4.5 is duplicated below, except the $\bar{C}_{km}$ values have been replaced by $E_{km}$ (the number of violations). Also shown in parentheses are the values of $\bar{E}_{km}$ ($= E_{km}/n_{km}$):

Table 4.7 Tabular Layout (similar to Table 4.5) for Weather Correcting the Number of Violation Days. ($n_k$ is the Total Number of Sample Days per Category, and $P_k^{(1)}$ is the A Priori Probability of Each Category.)

| | YEAR | | | | | | |
|---|---|---|---|---|---|---|---|
| CATEGORY | 1968 | 1969 | 1970 | 1971 | 1972 | $n_k$ | $P_k^{(1)}$ |
| 1 | 30*/30** (1.000) | 14/14 (1.000) | 12/12 (1.000) | 5/5 (1.000) | 18/17 (0.944) | 79 | .046 |
| 2 | 37/27 (0.730) | 21/12 (0.571) | 63/51 (0.809) | 20/15 (0.750) | 41/28 (0.683) | 188 | .106 |
| 3 | 103/5 (0.049) | 135/21 (0.155) | 118/13 (0.110) | 108/11 (0.102) | 115/12 (0.104) | 579 | .335 |
| 4 | 126/0 (0) | 97/1 (0) | 105/0 (0) | 161/0 (0) | 111/0 (0) | 600 | .349 |
| 5 | 61/0 (0) | 76/0 (0) | 41/0 (0) | 47/0 (0) | 55/0 (0) | 280 | .162 |

*Number of days in category.

**Number of days in category which violated standards.

The next step is to compute the expected frequency of violations per year under typical meteorological conditions. This is done by summing the product of the $P_k$ and $\bar{E}_{km}$ values, as follows:

Table 4.8 Normalization, by Category, of Violation Days (Same format as Table 4.5)

| CATEGORY | YEAR 1968 | 1969 | 1970 | 1971 | 1972 |
|---|---|---|---|---|---|
| 1 | .046 | .046 | .046 | .046 | .043 |
| 2 | .077 | .061 | .086 | .080 | .072 |
| 3 | .016 | .052 | .037 | .034 | .035 |
| 4 | .000 | .003 | .000 | .000 | .000 |
| 5 | .000 | .000 | .000 | .000 | .000 |
| Normalized frequency of violation days, $\hat{E}_m$ | .139 | .162 | .160 | .160 | .150 |
| Normalized number of days exceeding $SO_2$ standards ($E_m \cdot 365$) | 50.7 | 59.1 | 61.7 | 58.4 | 54.8 |

The final step is now to calculate the number of days per "typical" year that violations occurred. Resultant frequencies are multiplied by the number of days per year. Final results are shown in the bottom line of Table 4.8.

Whereas the actual data indicate significant variation in year-to-year exceedences of the $SO_2$ standards, the normalized values clearly indicate that conditions worsened through 1970, after which a noticeable improvement of violation-days occurred.

In summary, the recommended procedures offer a flexible approach for meteorologically adjusting both pollutant concentration levels and specified number of violation days. The most difficult aspect of the entire methodology will be the determination of appropriate meteorological classes. However, our recommended procedures, based on experience and empirical techniques, can assist in achieving an effective classification system. Once classification is accomplished, the "typical-year" meteorology approach can be applied without elaborate computation.

## REFERENCES


1. Bay Area Air Pollution Control District, "A Study of Oxidant Concentration Trends in the BAAPCD (1962-1972) Based on Temperature and Inversion Criteria," January 1973.

2. D. L. Blumenthal, et al., "Three Dimensional Pollutant Gradient Study--1972 Program," Meteorology Research, Inc., Altadena, May 18, 1973.

3. S. M. Bruntz, W. S. Cleveland, T. E. Graedel and B. Kleiner, Science, Vol. 186, 257 (1974).

4. _____, W. S. Cleveland, B. Keliner and J. L. Warner, "The Dependence of Ambient Ozone on Solar Radiation, Wind, Temperature, and Mixing Height," Preprint Volume, Amer. Meteor. Soc. Symposium on Atmospheric Diffusion and Air Pollution, 1974.

5. California Air Resources Board, (CARB) California Air Quality Data (CAQD), Vol. IV, No. 1, pp. 3-5.

6. _____, Vol. IV, No. 2, pp. 3-5.

7. _____, Vol. IV, No. 3, pp. 4-7.

8. _____, "Changes and Trends in Carbon Monoxide Levels," Vol. VII, No. 3, pp. 2-4.

9. _____, "Effect of Wind Speed on Pollutant Concentration," Vol. VI, No. 4, pp. 11-12.

10. _____, "Historical Occurrence of Temperatures Aloft Some 5000 Feet Above Sea Level," Vol. VII, No. 4, pp. 10-12.

11. _____, "Impact of Meteorological Conditions on Particulate Concentrations," Vol. VII, No. 4, pp. 13-14.

12. _____, "Oxidant Prediction," Vol. VI, No. 4, pp. 5-8.

13. _____, "Weekday vs. Weekend Oxidant Concentrations-II," Vol. VI, No. 4, pp. 10-11.

14. A. Davidson, "An Objective Ozone Forecast System for July Through October in the Los Angeles Basin," Los Angeles Air Pollution Control District, Technical Services Division Report, 1974.

15. A. Davidson, et al., "Air Quality Trends in Los Angeles County," Southern California Air Pollution Control District, Metropolitan Zone, December 1975.

16. Enviro Control, Inc., "Study in Trends in National Air Pollution and Related Effects, Phase I," prepared under Contract No. 68-01-0471 to the Environmental Protection Agency, May 1973.

17. K. E. Gronskei, E. Joranger and F. Gram, "Assessment of Air Quality in Oslo, Norway," Published as Appendix D to the NATO/CCMS Air Pollution Document "Guidelines to Assessment of Air Quality (Revised) $SO_x$, TSP, CO, HC, $NO_x$, Oxidants," Norwegian Institute for Air Research, Kjeller, Norway, February 1973.

18. E. L. Fisher, et al., "Forecasting Pollutant Dosages," Los Angeles County Air Pollution Control District, prepared for the California Air Resources Board under Standard Agreement ARB-1112, June 1973.

19. P. H. Guldberg, et al., "Air Quality Impact of the Energy Shortage," J. Appl. Meteor., 16, January 1977.

20. Y. Horie, and J. Trijonis, "Analysis and Interpretation of Trends in Air Quality and Population Exposure in the Los Angeles Basin," TSC-PD-A152-1, Technology Service Corporation, Santa Monica, Calif., March 1977 (to U.S. Environmental Protection Agency, Research Triangle Park, N.C.).

21. _____, and A. S. Chaplin, "Population Exposure to Oxidants and Nitrogen Dioxide in Los Angeles, Volume III: Long Term Trends, 1965-1974," U.S. EPA/OAQPS Publication #EPA-450/3-77-004c, Technology Service Corporation, Santa Monica, CA, January 1977.

22. R. M. Ingels, "Trends in Atmospheric Concentration of Oxides of Nitrogen 1957-1961," Los Angeles County Air Pollution Control District, August 1962.

23. T. R. Karl and G. A. DeMarrais, "Meteorological Conditions Conducive to High Levels of Ozone," International Conference on Photochemical Oxidant Pollution and Its Control, EPA-600/3-77-001a, January 1977.

24. R. E. Kerr, Jr., "Effects of Weather on Air Quality Measurements: San Gabriel Valley Ozone - Meteorological Model," Working Paper, June 1974.

25. J.J.R. Kinney, "Predicting Oxidant Concentrations in the Los Angeles Basin," California Air Resources Board, March 1974.

26. J. R. Kinosian and S. Duckworth, "Oxidant Trends in the South Coast Air Basin 1963-1972," California Air Resources Board, April 1973.

27. M. H. McCutchan and M. J. Schroeder, "Classification of Meteorological Patterns in Southern California by Discriminant Analysis," Journal of Applied Meteorology, Vol. 12, 571 (1972).

28. W. S. Meisel, "Empirical Techniques for Analyzing Air Quality and Meteorological Data. Part I: The Role of Empirical Methods in Air Quality and Meteorological Analyses," U.S. Environmental Protection Agency Publication #EPA-600/4-76-029a, Technology Service Corporation, Santa Monica, CA, July 1976.

29. M. K. Liu, et al., "Effects of Atmospheric Parameters on the Concentration of Photochemical Air Pollutants," J. Appl. Meteor., 15 (August 1976).

30. J. T. Peterson and E. C. Flowers, "Interactions between Air Pollution and Solar Radiation," Solar Energy, Vol. 19, 1977.

31. J. H. Price, "A Study of Factors Associated with High Urban Ozone Concentrations in Texas," Texas Air Control Board, Presented at the Air Pollution Control Association Conference on Ozone/Oxidants, Dallas, Texas, March 1976.

32. E. A. Schuck, J. N. Pitts, Jr., and J.K.S. Wan, Air & Wat. Pollut. Int. J., Vol. 10, 689 (1966).

33. S. M. Sidik, and H. E. Neustadter, "Meteorological Adjustment of Yearly Mean Values for Air Pollutant Concentration Comparisons," NASA Technical Note TN D-8253, National Aeronautics and Space Administration, Washington, D.C., July 1976.

34. J. A. Sonquist, E. L. Baker, J. N. Morgan, Searching for Structure, University of Michigan, 1971.

35. G. C. Tiao, G.E.P. Box, and W. J. Hamming, "A Statistical Analysis of the Los Angeles Ambient Carbon Monoxide Data 1955-1972," J. Air Poll. Control Assoc., Vol. 25 (Nov. 1975).

36. _____, M. S. Phadke, and G.E.P. Box, "Some Empirical Models for the Los Angeles Photochemical Smog Data," J. Air Poll. Control Assoc., Vol. 26 (May 1976).

37. J. C. Trijonis, "An Economic Air Pollution Control Model - Application: Photochemical Smog in Los Angeles County in 1975," Ph.D. Thesis, California Institute of Technology, May 1972.

38. _____, et al., "Emissions and Air Quality Trends in the South Coast Air Basin," EQL Memo No. 16, Environmental Quality Laboratory, California Institute of Technology, Pasadena, CA , January 1976.

39. _____, "Empirical Studies of Nitrogen Dioxide Air Quality and $NO_2$/Precursor Relationships," Report in preparation to the U.S. Environmental Protection Agency under Contract No. TSC 68-02-2299.

40. U.S. Environmental Protection Agency, "Monitoring and Air Quality Trends Report, 1973," EPA-450/1-74-007, October 1974.

41. _____, "Monitoring and Air Quality Trends Report, 1974," EPA-450/1-76-001, February 1976.

42. _____, "National Air Quality and Emission Trends Report, 1975," EPA-450/1-76-002, November 1976.

43. H. van Dop, and S. Kruizinga, "The Decrease of Sulphur Dioxide Concentrations near Rotterdam and Their Relation to Some Meteorological Parameters During Thirteen Consecutive Winters (1961-1974)," Atmospheric Environment, Vol. 10, pg. 1 (1976).

44. M. D. Zeldin and D. Thomas, "Ozone Trends in the Eastern Los Angeles Basin Corrected for Meteorological Variations," presented at the International Conference on Environmental Sensing and Assessment, Las Vegas, Nevada, 1975.

45. _____, A. Davidson, M. F. Bronelle, and J. E. Dickinson, "A Meteorological Assessment of Ozone and Sulfate Concentrations in Southern California," Southern California Air Pollution Control District, Report #E&P 76-1, August 1976.

46. _____, A. M. Winer, D. Grosjean, and J. N. Pitts, Jr., "Weather Adjusted Oxidant Trends in the South Coast Air Basin," Statewide Air Pollution Research Center, Riverside, Calif., May 1977 (to be submitted to Atmospheric Environment).

47. California Air Resources Board, "Sulfate Predictions in South Coast Air Basin," Vol. VIII, No. 2, pp. 4-10.

48. L. Breiman, Statistics: With a View Toward Applications, Houghton-Mifflin, 1973.

49. H.J. Payne and W. S. Meisel, "An Algorithm for Constructing Optimal Binary Decision Trees," IEEE Trans. on Computers, Vol. C-26, No. 9, September 1977, pp. 905-916.

# APPENDIX A
## MULTIPLE-YEAR LINEAR REGRESSION

Suppose we wish to determine a linear equation relating a meteorological variable to a pollutant concentration. Suppose further that we had several years of data. Simply regressing the meteorological variable against the pollution concentration for all the data would not be appropriate. The coefficients of the meteorological variables might be misleading if emissions change from year to year. Figure A-1 exemplifies this problem. Suppose in two consecutive years, the pollutant level was predicted well by a meteorological variable x. Suppose the best-fit equations were

$$y = ax + b_1$$

and

$$y = a'x + b_2 \quad ,$$

respectively, where $a \approx a'$. Then, the change in y corresponding to a change in x in a given year is closely given by the coefficient a or a'. However, if there is a change in the overall level of emissions and different mean values of x, say $m_1$ and $m_2$, in the two years, the fit to the data of both years can have a considerably different slope. (See Figure A-1.) Thus, we cannot simply fit data over many years to meteorological variables alone.

Suppose we instead put in "dummy" variables indicating the year from which the data arose:

$$z_i = \begin{cases} 1 \text{ if the data is from year } i \\ 0 \text{ otherwise} \quad . \end{cases} \qquad \text{(A-1)}$$

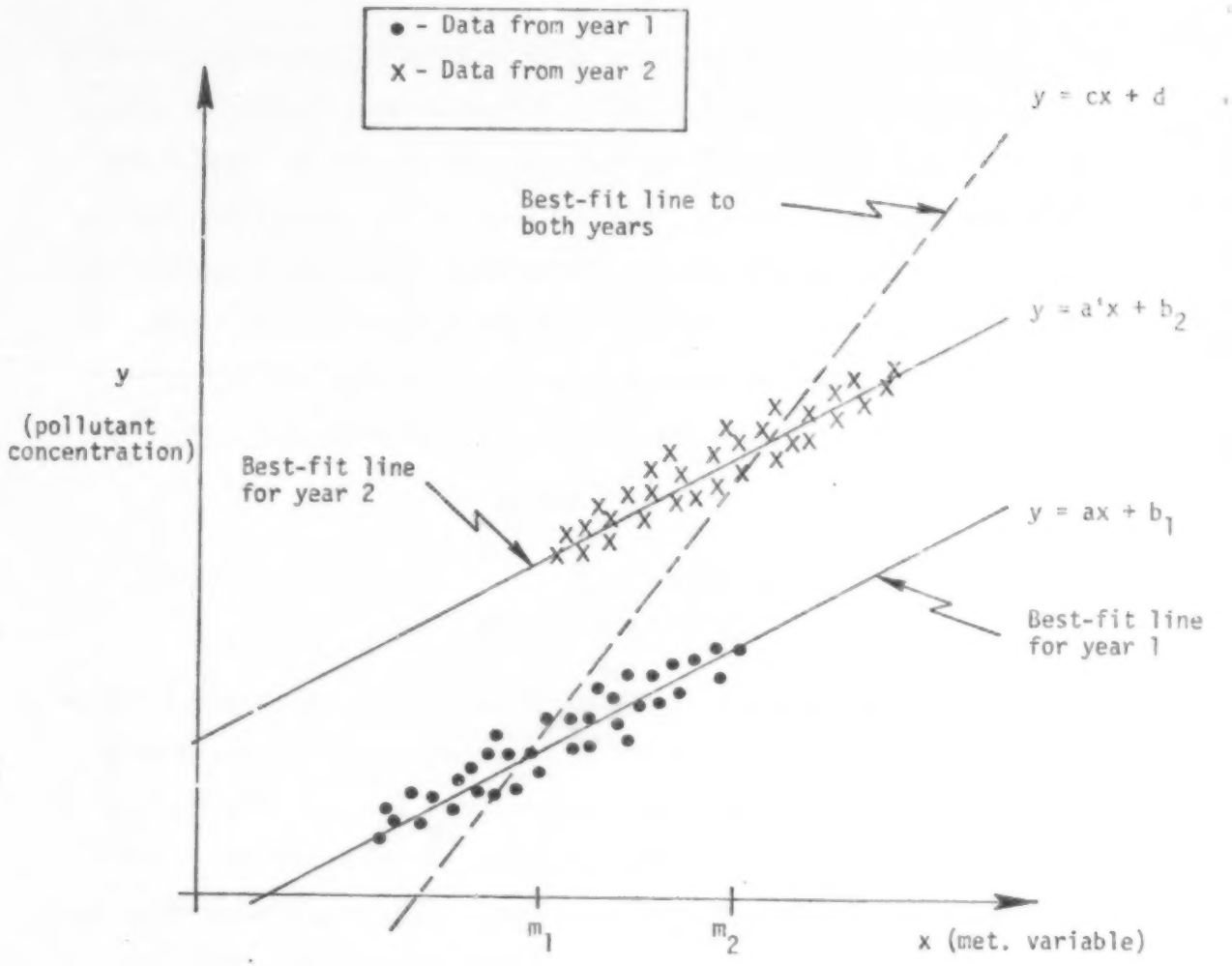


Figure A-1 An Example of the Difference in Fitting Each Year Separately Rather than Fitting All the Data.

Then we perform a regression of the form

$$y = ax + c + \sum_{i=1}^{K} b_i z_i \qquad \text{(A-2)}$$

for K years. The model we are assuming is then that there is a single slope for all years but a different constant term. The regression will give us the best coefficient "a" in a least-squares sense. This approach essentially allows us to derive two separate fits to data from the two years with the constraint that the coefficient of the meteorological variable is the same in each year.

Note that this procedure is not equivalent to using a single dummy variable assuming a linear trend. The assumption of a linear trend can distort the results if the true trend is, for example, U-shaped.

The generalization of this approach to N meteorological variables is the equation

$$y = \sum_{i=1}^{N} a_i x_i + c + \sum_{i=1}^{K} b_i z_i \quad . \qquad \text{(A-3)}$$

This assumes that the coefficient on each variable is the same for all K years.*

*Note that this is a regression of daily meteorological variables against daily pollution levels.

# APPENDIX B

## A METHODOLOGY FOR THE DEFINITION OF METEOROLOGICAL CLASSES

The procedure outlined in this appendix requires meteorological and pollution data from several years. The methodology can be employed using standard statistical software with human intervention. The steps are as follows:

*(A) Select Data*

The first step in the process is to assemble the meteorological and air quality data for some representative station with sufficient meteorological parameters. A period of at least 2 or 3 years should be used in the initial processes. The selection of the pollutant parameter should be related to some time-averaging period specified in the NAAQS (e.g., 1-hour maximum hourly average for oxidant, 1- or 8-hour averages for CO, daily values for TSP, etc.) or some equivalent.

*(B) Derive Initial Classes*

For each meteorological variable to be considered, plot the value of meteorological variable, x, versus the value of pollutant variable, C, for each day. By visual inspection, the meteorological variable which best allows a discrimination of pollutant concentration should be selected. Call this weather variable $x_1$. As shown in Figure B-1, we want to be able to find one or two regions where the pollutant level has a small variation about its average value (e.g., region $R_2$). This satisfies a condition whereby the meteorological parameter closely predicts the pollutant level; therefore,

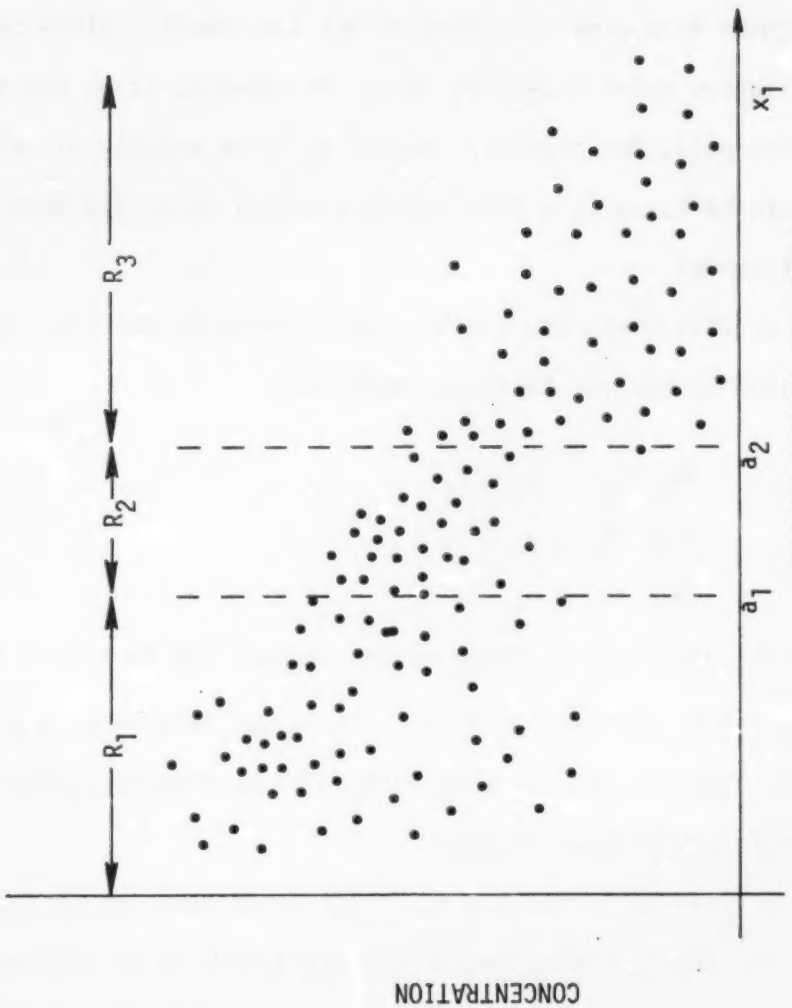


Figure B-1 Sample Scatterplot Defining Three Distinct Regions. (Regions $R_2$ and $R_3$ can be considered final classes; region $R_1$ requires further study.)

this region is a candidate for a final meteorological class. In this case, further examination is not merited because the pollutant prediction is sufficiently good. Such a class, however, must have sufficient independent points for the classification to be meaningful.

Another type of region to look for is one where no high pollutant levels occur, even if there is a wide variation at the low levels. Since it is the high-pollution classes where attention should be focused, it is not justified to proliferate low-pollution classes. Region $R_3$ is an example of this type of region. It can be taken as a final meteorological class and does not need further refinement.

At the end of this step, one meteorological variable has been chosen, two regions defined by the variable, $x_1$, such that

$$R_1: \quad x_1 \leq a_1$$
$$R_2: \quad a_1 < x_1 \leq a_2$$
$$R_3: \quad a_2 < x_1.$$

where $a_n$ denote the values of $x_1$ bounding the regions (as in Figure B-1). Specifically, $R_2$ and $R_3$ are final classes. It is not necessary to process them any further; however, regions such as $R_1$ must be treated further.

*(C) Refine Poorly Defined Regions*

For each undefined region (e.g., $R_1$), plot those cases which fall in that region against the remaining meteorological variables as before. Using the best meteorological variable for doing so, divide data from $R_1$ into two or three regions as before. Continuing the example, suppose $x_2$ was the best such variable (Figure B-2). $R_1$ is broken up by meteorological parameter $x_2$, into two regions:

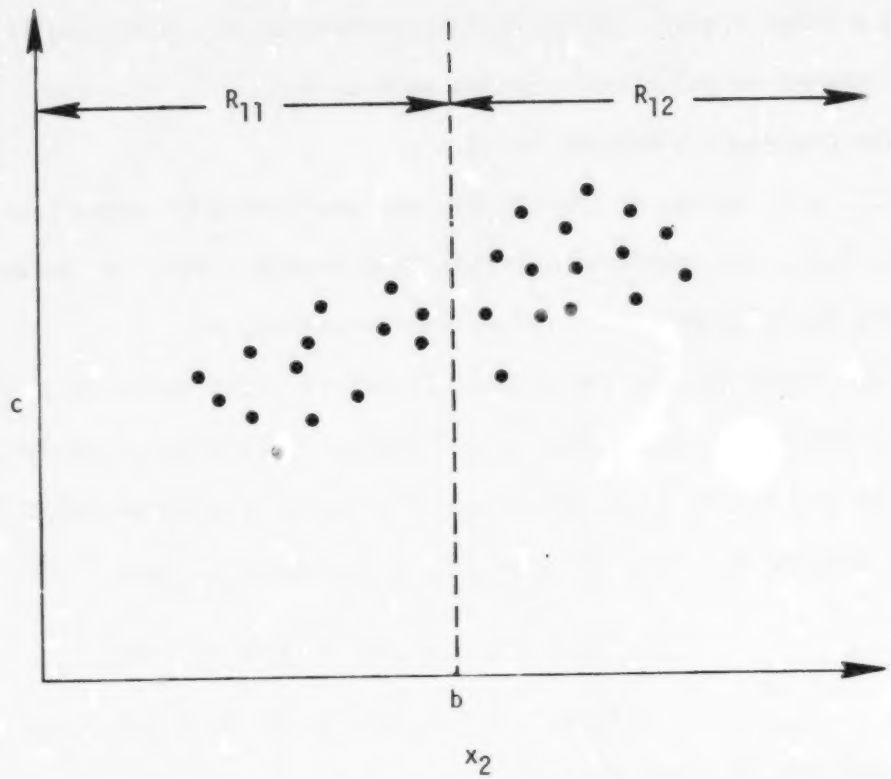


Figure B-2 Data from Region $R_1$ (i.e., for which $x_1 \leq a_1$) Plotted for Another Meteorological Variable, $x_2$.

$$R_{11}: \quad x_2 \leq b \text{ [and } x_1 \leq a_1\text{]}$$

$$R_{12}: \quad x_2 > b \text{ [and } x_1 \leq a_1\text{]} \quad .$$

$R_{12}$ is a final region. Continue this process on any remaining ill-defined (with respect to pollutant) regions such as $R_{11}$.

The procedure continues until

(1) all regions are relatively well-defined with respect to pollutant levels, i.e., the standard deviation in pollutant levels is low or there are very few high-pollution values in the region; or

(2) there are too few samples in an uncertain region to divide it further and retain statistical significance. (Formally, a statistical test could be applied for statistical significance of a proposed split [48].)

Suppose $R_{11}$ is split once more by parameter $x_3$ into

$$R_{111}: \quad x_3 \leq c \text{ [and } x_2 \leq b, \; x_1 \leq a_1\text{]}$$

$$R_{112}: \quad x_3 > c \text{ [and } x_2 \leq b, \; x_1 \leq a_1\text{]}$$

and these are all final classes.

The result is a decision tree defining the meteorological categories as illustrated in Figure B-3.

Extensions

We have used one variable at each step. As in the example used in Section 4.1.2 of the text, we could use two (or more) variables in any step. The ideas are easy to extend, but the software (e.g., University of Michigan's AID and TSC's ADSS program) required is less standard.

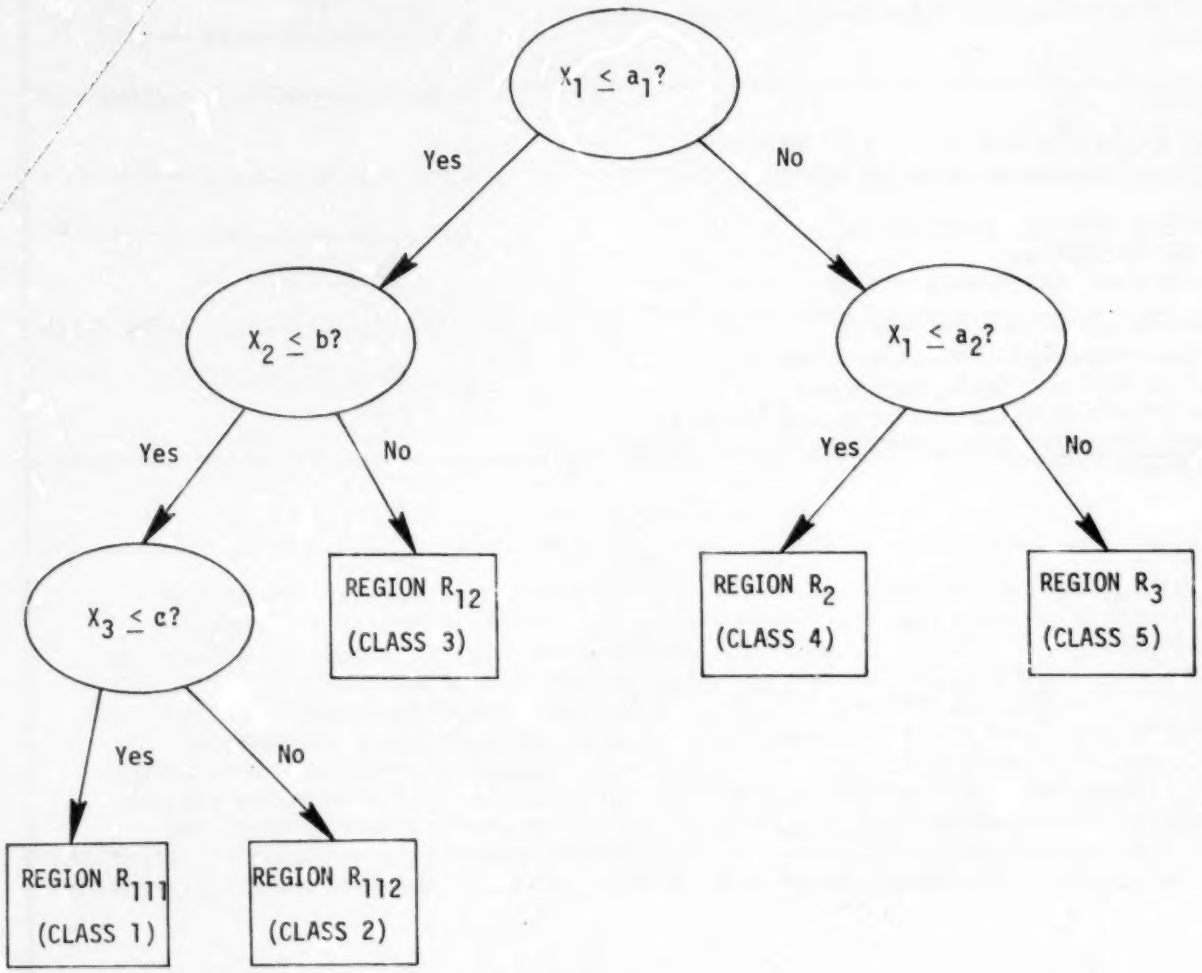


Figure B-3 Decision Tree Approach Toward Defining Meteorological Classes Based on Meteorological Parameters $X_1$, $X_2$, and $X_3$.